The Ice Cloud Imager: retrieval of frozen water column properties

Eleanor May[1], Bengt Rydberg[2], Inderpreet Kaur[1,2], Vinia Mattioli[3], Hanna Hallborn[1], and Patrick Eriksson[1]

[1]Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden
[2]Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden
[3]EUMETSAT, Darmstadt, Germany

Correspondence: Eleanor May (eleanor.may@chalmers.se)

Abstract.

The Ice Cloud Imager (ICI) aboard the Second Generation of the EUMETSAT Polar System (EPS-SG) will provide novel measurements of ice hydrometeors. ICI is a passive conically scanning radiometer that will operate within a frequency range of 183 GHz to 664 GHz, helping to cover the present wavelength gap between microwave and infrared observations. Reliable

5   global data will be produced on a daily basis. This paper presents the retrieval database to be used operationally and performs a final pre-launch assessment of ICI retrievals.

Simulations are performed within atmospheric states that are consistent with radar reflectivities and represent the three-dimensional variability of clouds. The radiative transfer calculations use empirically-based hydrometeor models. Azimuthal orientation of particles is mimicked, allowing for polarisation to be considered. The degrees of freedom of the ICI retrieval

10   database are shown to vary according to cloud type. The simulations are considered to be the most detailed performed to this date. Simulated radiances are shown to be statistically consistent with real observations.

Machine learning is applied to perform inversions of the simulated ICI observations. The method used allows for the estimation of non-Gaussian uncertainties for each retrieved case. Retrievals of ice water path (IWP), mean mass height (Z_m) and mean mass diameter (D_m) are presented. Distributions and zonal means of both database and retrieved IWP show agreement

15   with DARDAR. Retrieval tests indicate that ICI will be sensitive to IWP between 10^{-2} and 10^{1} kg m^{-2}. Retrieval performance is shown to vary with climatic region and surface type, with the best performance achieved over tropical regions and over ocean. As a consequence of this study, retrievals on real observations will be possible from day one of the ICI operational phase.

1   Introduction

20   There is an urgent need for reliable and consistent estimates of ice mass in the atmosphere. Measurements of ice mass are pivotal to understanding global weather patterns and current climate trends. The release of latent heat upon ice formation drives circulation within the atmospheric water cycle (Bony et al., 2015), giving rise to events such as deep convective systems. Frozen hydrometeors also constitute the ice clouds in our atmosphere. Such clouds modulate the amount of outgoing longwave

radiation (OLR) and reflect solar radiation, and thus hold significant influence over the climate. Ice clouds are now considered
a major climate feedback (Stephens et al., 1990).

Despite their importance, the mass of ice hydrometeors in the atmosphere remains an uncertain quantity. A significant factor
contributing to this uncertainty is the complex relationship between ice mass and satellite observation — the relationship is
difficult to characterise, thus leading to a large spread between models and data-derived observational products, and even
between the products themselves. Duncan and Eriksson (2018) highlighted these shortcomings by showing the significant
discrepancies that exist between cloud ice datasets. Consequently, ice hydrometeor mass observations remain a gap in the
atmospheric hydrological cycle. Given the present-day state of the climate, it is becoming increasingly necessary to address
this gap (Waliser et al., 2009).

The upcoming Ice Cloud Imager (ICI), as part of the EUMETSAT Polar System - Second Generation (EPS-SG) programme,
will address this issue by specifically measuring atmospheric ice mass at microwave and sub-millimetre wavelengths (Mattioli
et al., 2019). While current satellite observations span a wide range of frequencies, many face limitations in the measurement
of ice hydrometeors. Passive optical and infrared missions are sensitive to small ice hydrometeors, but predominantly capture
only cloud top data due to high attenuation. Passive microwave sensors are able to penetrate cloud, but measure at wavelengths
that are sensitive only to the largest ice hydrometeors. High-frequency radar has good sensitivity to ice, but are limited to
observe at only a single frequency and thus poorly constrain hydrometeor properties. No operational sensors currently provide
passive sub-millimetre observations of our atmosphere. However, ICI will pair sub-millimetre wavelength observations with
passive microwave observations. This allows for sensitivity to a wider range of ice hydrometeor sizes, and increases the overall
information content through the use of multiple frequencies. ICI will cover a frequency range of 183 GHZ to 664 GHz.
Alongside ICI on the same satellite will be the MWI (Micro-Wave Imager), extending the coverage to lower frequencies.
Together, ICI and MWI will offer an unparalleled collection of passive measurements, closing the existing frequency gap in
ice hydrometeor mass observations (Accadia et al., 2020).

Evidence supports the improvement of cloud ice retrievals in the presence of sub-millimetre observations. Successful cloud
ice retrievals have been performed using sub-millimetre limb sounder observations (Wu et al., 2008; Eriksson et al., 2014).
Brath et al. (2018) demonstrated that retrievals of ice water path (IWP) improved when combining microwave and sub-
millimetre flight campaign observations. Pfreundschuh et al. (2022b) showed enhanced sensitivity to hydrometeors when
supplementing radar data with sub-millimetre observations. In the context of a future ICI launch, Wang et al. (2016) per-
formed successful retrievals of cloud parameters based on a synthetic database of ICI radiances. Finally, Eriksson et al. (2020)
— the precursor to this study — presented the operational algorithm to be used within the ICI Level 2 (L2) product at EU-
METSAT. In the study, a preliminary database of simulated ICI observations was used to demonstrate successful ICI retrievals
and to estimate the retrieval performance.

This study presents the generation of the final retrieval database, composed of ICI observation simulations and corresponding
ice mass quantities. This database serves to complete the operational algorithm, and performs a pre-launch characterisation of
ICI retrievals. We emulate retrieval performance using synthetic ICI data, where retrievals meet operational standards, i.e. only
ice hydrometeors are considered. The study focuses only on channels present on ICI, and MWI measurement frequencies are

not considered. The primary aim of the study is to confirm that both simulations and retrievals are statistically consistent with reality. In the absence of actual 'ground truth', flight-campaign data and existing retrieval products are used. Additionally, we investigate how retrieval performance varies according to climatic region, surface conditions, and noise estimates. In summary, we perform the best possible pre-launch assessment of the retrieval of ice hydrometeor column values. To this end, we evaluate whether the ICI L2 product is set to deliver reliable information on ice mass in the form of ice water path (IWP), mean mass height (Z_m) and mean mass diameter (D_m), and under which conditions.

A description of the retrieval database and the database generation techniques are given in Sect. 3. Details on the retrieval approach are given in Sect. 4. In Sect. 5, radiances resulting from simulations of two similar instruments are compared with real observations, and an assessment of the overall retrieval performance is given.

2 The Ice Cloud Imager (ICI) Mission

2.1 The ICI instrument

ICI will be hosted on the Metop-SG B satellite series as part of the EPS-SG mission. Each satellite has a planned lifetime of 7.5 years. With three successive launches of a pair of satellites — Metop-SG A and B — continuous coverage for over 21 years will be achieved. The satellites will be in a sun-synchronous orbit and fly at a height within the range of 823 - 843 km. At the time of writing, the launch of the first satellite in each of the two EPS-SG series is scheduled for 2025 and 2026, respectively.

The ICI instrument is a conically scanning radiometer, providing observations at an incidence angle of $53° \pm 2°$. Observations are taken over an angle of $\pm 65°$ around the forward view of the orbital path, giving a swath width of approximately 1700 km, which results in near-global coverage on a daily basis. ICI will observe using 13 channels, operating at local oscillator frequencies of 183.31, 243.2, 325.15, 448.0, and 664.0 GHz. Specifications of the channels can be found in Eriksson et al. (2020). Nine channels around 183.31, 325.15 and 448.0 GHz cover water vapour molecule transitions. The remaining channels at 243.2 and 664.0 GHz function as 'window' channels, where decreased absorption by atmospheric gases enables observations down to relatively low altitudes. The window channels will measure at both horizontal and vertical polarisation. As such, ICI has the potential to capture the effects of oriented hydrometeors and of polarisation as a result of surface interaction. The other channels (183.31, 325.15 and 448.0 GHz) will measure at only vertical polarisation. Further details on the receivers, the antenna system and calibration are given in Eriksson et al. (2020). However, some updates have been made since the publication of Eriksson et al. (2020); detailed estimates of the spectral response function and the antenna response function are now available. Furthermore, first actual estimates of the noise performance are now at hand, indicating that the noise requirements will be met with a margin. Consequently, we assume below that the magnitude of thermal noise is at 75% of the NEΔT estimates given in Eriksson et al. (2020).

Also on board Metop-SG B will be MWI — another conically scanning radiometer that will also observe at an incidence angle of $\approx 53°$. MWI will measure at frequencies between 18.7 GHz and 183 GHz, allowing for sensitivity to precipitation and snowfall. It will offer both horizontal and vertical polarisation up to frequencies of 89 GHz, and vertical polarisation only for the higher-frequency channels.

2.2 Operational L2 product

EUMETSAT will offer a Level 2 product (MWI-ICI-L2) containing retrievals based on ICI and MWI observations in near real-time. Although retrievals will be performed separately for ICI and MWI using different retrieval methods, they will be present in the same product. This is the only operational L2 product planned. As a result of the MWI contribution, MWI-ICI L2 will offer liquid water path (LWP). The primary variable offered on the ICI-side of the MWI-ICI L2 product is ice water path (IWP), defined as the vertical column of atmospheric ice water content. Mean mass height (Z_m) and mean mass diameter (D_m) are also offered as retrieval products for ICI, existing only for IWP > 0 kg m^{-2}. This study deals only with the ICI-side variables, of which definitions are given in Eriksson et al. (2020). The algorithm theoretical basis document (ATBD) is provided by the EUMETSAT Satellite Application Facility (SAF) supporting nowcasting (NWC-SAF) (Rydberg, 2018). The algorithm is finalised. What remains for the L2 product is the development of a database (discussed below), which is of major importance for retrieval accuracy.

Within the algorithm, pre-processing is undertaken before the inversion takes place. The pre-processing steps include the remapping of L1b data to a common footprint, that was the output of an optimal interpolation scheme developed within a EUMETSAT study with results presented in Eriksson et al. (2020). Additionally, there is a bias correction scheme applied in the case of any systematic differences between L1b data and retrieval database cases. An optional module for the detection of clear-sky cases, i.e. IWP = 0, is also included within the algorithm. The reader is referred to Eriksson et al. (2020) for further details of the pre-processing steps.

To perform the inversion, the retrieval algorithm uses an implementation of Bayesian Monte Carlo integration (BMCI) (Eriksson et al., 2020). The inversion itself is centred around a retrieval database — a dataset containing ice mass products and associated observations. The aim of the algorithm is to map observations to ice mass products, using the retrieval database as reference data. The input to the algorithm is a vector of measurements, and the outputs are the retrieval products (IWP, Z_m, D_m) associated with the given measurements. The measurements take the form of geolocated and calibrated antenna temperatures. Alternatively, the measurement vector can be replaced with a cloud signal, i.e. a simulated clear-sky observation subtracted from the original all-sky case. The ICI retrieval algorithm takes the latter approach. This has the advantage of reducing the background contribution to the observation, where Rydberg (2018) can be referred to for more detail.

In Rydberg (2018), a retrieval database was developed with the aim of testing the ICI retrieval algorithm and estimating retrieval performance. Although the simulations were detailed, this preliminary database was limited in some aspects. Particle orientation effects and three-dimensional variability within clouds were not accounted for. The purpose of this study is the development of a new and final retrieval database for operational use in the L2 product and an updated characterisation of expected retrievals. We refer to the database of this previous study as the *preliminary database*. The database generation method has been updated to reflect any recent developments, and the limitations of the preliminary database are addressed. This is discussed further in the following section. The database generated in this study is referred to as the *new database* or, simply, the retrieval database. The development of the operational L2 product at EUMETSAT was finalised with the completion of this new retrieval database.

2.3 Numerical weather prediction

The main application of ICI is to provide global measurements related to ice clouds in support of climate monitoring. Another application area of ICI is numerical weather prediction (NWP). ICI could be of some interest for assimilation systems of clear-sky character; its radiances would provide some additional information on humidity, or could be used for cloud filtering

130 (Kaur et al., 2021) of MWI data. However, as ICI is designed to constrain ice hydrometeor properties (Buehler et al., 2007), its data will be most useful if used in all-sky assimilation. As a consequence, the preparatory work inside NWP and developing L2 products has considerable overlap, both applications require development of radiative transfer involving scattering of ice particles.

A common step in this direction was the EUMETSAT study that resulted in the single-scattering data presented in Eriks-

135 son et al. (2018). This "ARTS database" is not only used for the L2 retrievals of concern here, but parts of it are now also incorporated in the leading NWP forward operators RTTOV (Geer et al., 2021) and the Community Radiative Transfer Model (CRTM) (Moradi et al., 2022). One aim of Eriksson et al. (2018) was to offer a broad set of particle shapes, and for practical issues effects of particle orientation could not be accommodated in this relatively large database. As a complement, Brath et al. (2020) covers particle orientation for two habits. The later data were developed for supporting research and L2 retrievals, but

140 turned out to be vital for developing a fast scheme to approximate the impact of particle orientation (Barlakas et al., 2021) inside RTTOV-SCATT (the microwave all-sky module of RTTOV). This development, in its turn, triggered an extension of the fast scheme that became used in the production of the new retrieval database (Sect. 3).

The ARTS software itself (Buehler et al., 2018) is supporting NWP by acting as a reference model (Barlakas et al., 2022). Reversely, the new microwave absorption model developed for RTTOV is applied when running ARTS inside this study (Sect.

145 3.1). Largely based on these efforts, there is already infrastructure in place at ECMWF to make use of ICI radiances in all-sky assimilation (Alan Geer, ECMWF, private communication).

3 Retrieval database

The novelty of ICI lies in its ability to provide new and unique observations at sub-millimetre wavelengths. However, this presents a challenge when creating a retrieval database, given the current unavailability of sub-millimetre observations. The

150 retrieval database must be assembled in the absence of actual observations, preventing the creation of an empirically-based database. Instead, it is necessary to simulate ICI observations.

There have been several attempts at generating retrieval databases of simulated sub-millimetre observations. One early use of a retrieval database is Evans et al. (2002). Stochastic cloud profiles were constructed based on the in-situ microphysical data, which are then used within radiative transfer calculations. The result was simulated brightness temperatures as they

155 would be measured by the aircraft-mounted Submillimetre-Wave Cloud Ice Radiometer (SWCIR). A later attempt to develop a retrieval database is detailed in Rydberg et al. (2007). This database was limited to one-dimensional atmospheric states, used only single frequency channels, and was limited to northern mid-latitudes. The method was later extended (Rydberg et al., 2009) to include CloudSat data in the generation of three-dimensional atmospheric scenes using a Fourier transform algorithm,

but simplified cloud microphysics remained a limitation. Evans et al. (2012) performed IWC retrievals using sub-millimetre
flight campaign data. The retrieval database was built upon stochastic hydrometeor profiles that were generated using CloudSat
radar reflectivities and microphysics probability distributions. The microphysical information was based on in-situ data, and is
therefore detailed but limited by the region of flight and the location of in-cloud sampling. In Wang et al. (2016), a synthetic
database of ICI observations was created by performing radiative transfer calculations on simulated hydrometeor profiles.
The database was successfully used to perform cloud ice retrievals specifically for ICI, but was limited to cases over Europe.
Finally, the preliminary database developed by Eriksson et al. (2020) marked the most recent major development. The database
reflects global variability with regard to atmospheric and surface scenarios. Additionally, the microphysical assumptions were
improved. However, some areas remained to be improved upon for the new retrieval database.

3.1 Overview

The reliability of the retrievals relies strongly on the quality of the retrieval database. As such, there are several requirements
placed on the database. Firstly, the database must contain realistic simulations of the future ICI observations, thus considering
all relevant atmospheric, surface and instrumental variables. Secondly, the database as a whole shall statistically represent the
variability of the observations. Finally, to ensure successful inversions, there should be a sufficiently large number of database
cases to ensure that each observation matches multiple states. Therefore, a high number of rigorous all-sky simulations must
be performed.

The basic strategy applied for the generation of the preliminary database was found consistent with the requirements and
was therefore kept. However, several extensions were required for the new database. Therefore, a full reimplementation of
the simulation environment was necessary in order to increase the calculation efficiency in such way that the extensions were
feasible. In particular, a simplified description of the sensor was used for the generation of the preliminary database. Incorpo-
rating a more complete description of the antenna and channel responses in the new database results in a significant increase
in radiative transfer calculations.

A second main area of improvement was to extend the description of ice hydrometeor properties to include a non-uniform
probability of occurrence and a computationally efficient, approximate treatment of particle orientation. The latter was required
to cover the polarisation response of ICI's channels, which was not accounted for in the preliminary database. Remaining
improvements include considering the full interference of ozone in the atmospheric radiative transfer calculations, a better
description of the emissivity of snow-covered surfaces, and a small correction originating in the calibration procedure.

The omission of a full antenna pattern has been a significant limitation in all previous attempts to develop retrieval databases.
If an observation is represented with a single pencil beam calculation, the decrease in radiance caused by the presence of ice
hydrometeors tends to be overestimated (Barlakas and Eriksson, 2020). Accordingly, the 'beam-filling' must be captured
(Davis et al., 2007). By including the along-track antenna response, the preliminary database went further than most earlier
works. However, incorporating a two-dimensional antenna response provided the largest challenge in the extension of the
framework. This challenge had several parts. Firstly, it raised the question of how to generate three-dimensional atmospheric
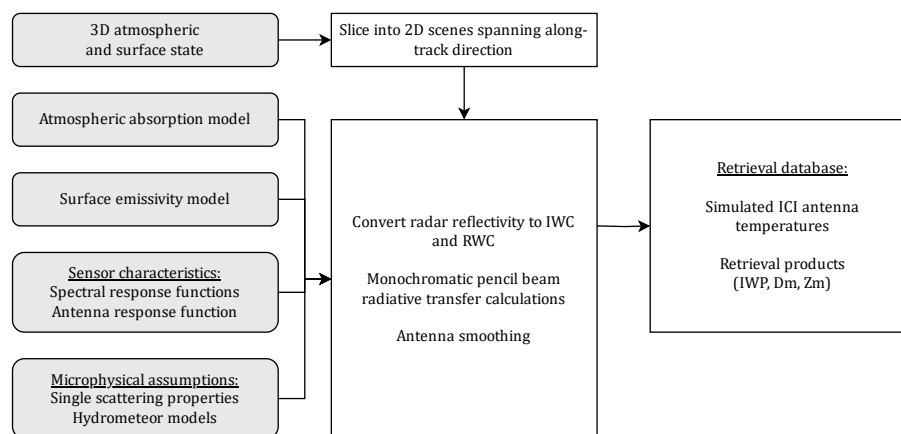
Figure 1. Summary of the inputs, processes and outputs of the retrieval database generation scheme.

scenes based on CloudSat data that have only a coverage of two-dimensional character (height and along-track). This was solved by implementing the method of Barker et al. (2011).

The next consideration is how to perform the radiative transfer calculations. Full three-dimensional calculations are infeasible due to the associated high computational burden. Instead, an independent beam approximation (IBA) was selected. This entails sampling the atmosphere along a number of (slanted) directions and performing local one-dimensional radiative calculations. This is followed by a weighting of the obtained pencil beam brightness temperatures with the antenna response to obtain the antenna temperature for each frequency. Barlakas and Eriksson (2020) showed that this IBA approach, when applied within the frequency range of ICI, removes the simulation bias typically present for single pencil beam calculations, albeit with a smaller random error remaining. More details are found in Sect. 3.5.

A summary of the inputs, process and outputs of the simulation environment is given in Fig. 1. The different parts of the framework are presented in the subsequent sections. For further details, the reader may refer to the study report (Rydberg et al., 2023). The content of the database delivered to EUMETSAT is found in Table 1. The core variables of the database are IWP, Z_m and D_m, and corresponding simulated ICI observations in the form of antenna temperatures T_a. The atmospheric variables are antenna-weighted values and there is an assumption of remapped ICI data to a common footprint (see Eriksson et al. (2020)). An extended database containing, for example, the profiles of ice water content underlying the IWP values, has been saved to be used for future research purposes.

3.2 Generation of 3D hydrometeor fields

Cloud radars are currently the best source of information on vertical cloud structure. The 94 GHz Cloud Profiling Radar (CPR) aboard CloudSat provides measurements in the form of radar reflectivities at 500 m vertical resolution, offering nearly global coverage. CloudSat data from 2009 and 2010 were used, avoiding the fact that CloudSat offers only daytime coverage since 2011. However, CloudSat offers only two-dimensional data, lacking information in the across-track direction. To expand

Table 1. Variables included in the ICI cloud radiation retrieval database to be used operationally at EUMETSAT. Ice water path refers to the *total* ice water path, i.e. both cloud ice water path and precipitating ice.

Variable	Unit	Description
$T_{a,cs,i}$	K	Simulated clear sky antenna temperature for ICI channel i.
$T_{a,as,i}$	K	Simulated all sky antenna temperature for ICI channel i.
$\Delta T_{a,i}$	K	Simulated cloud signals in antenna temperature (Eq. 3).
τ_i	-	Cloud optical depth for ICI channel i.
IWP	kg m^{-2}	Ice water path (total).
Z_m	m	Mean ice hydrometeor height by mass.
D_m	m	Mean particle size by mass, in terms of a mass equivalent spherical diameter.
RWP	kg m^{-2}	Rain water path.
LWP	kg m^{-2}	Liquid water path.
TCWV	kg m^{-2}	Column water vapour.
Surface Type	-	Possible types: ocean, land, snow, sea ice, mixed.
Surface pressure	Pa	-
Surface temperature	K	-
Surface wind	m s^{-1}	Surface wind speed.
Latitude	degrees	-
Longitude	degrees	-
Ice habit	-	Frozen hydrometeor model, incorporating habit and PSD.

into a three-dimensional representation, and therefore cover horizontal cloud structure, the 3D cloud-construction algorithm of Barker et al. (2011) was implemented. In the algorithm, a three-dimensional cloud structure is generated through the colocation

215 of passive satellite and two-dimensional radar data. Both CloudSat and The Moderate Resolution Imaging Spectroradiometer (MODIS) onboard Aqua are part of NASA'S A-train constellation, and therefore collocated data from the two instruments exist. The Barker algorithm was applied directly to CloudSat radar reflectivities, avoiding the use of retrieved data.

The three-dimensional cloud structures obtained using the Barker algorithm are then merged with ERA5 data to provide a full description of atmospheric gases, meteorological conditions and the surface type within the scene. The output is a three-

220 dimensional scene containing all relevant atmospheric and surface parameters.

Finally, the three-dimensional distributions of radar reflectivities are converted to fields of hydrometeor contents. The relationship between radar reflectivity and IWC is conditioned on microphysical assumptions (Kulie et al., 2010; Ekelund et al., 2020). The reflectivities, based on the microphysics described in Sect. 3.3, are calculated on a grid of IWC, leading to a table of IWC-dBZ values and associated attenuation. The inversion of radar reflectivities within a given scene to IWC fields makes

225 use of the look-up table, following the setup described in Sect. 3.2 of Ekelund et al. (2020). The same approach is used for the

Table 2. Particle models used within simulations of the ICI retrieval database, given alongside habit and particle size distribution (PSD). The source for the PSD is given in each case. The aARO scaling factor is randomly chosen within the given range. Also given is the occurrence fraction, or the probability that a given model is randomly selected for each atmospheric scene, denoted as p_i

Model	Habit	PSD	aARO factor	p_i
AA1	Large plate aggregate	F07 - Tropics	1 - 1.6	0.3
AA2	Large column aggregate	F07 - Tropics	1 - 1.6	0.1
AA3	Large block aggregate	D14	1 - 1.6	0.13
IWC	Six bullet rosette	D14	1 - 1.6	0.2
Snow	Evans snow aggregate	F07 - Midlatitude	1.4 - 1.6	0.1
Graupel	Eight column aggregate	D14	1 - 1.2	0.17

generation of rain water content for temperatures above $0°C$. Liquid water content (LWC) is taken from ERA5, since CloudSat is largely insensitive to non-precipitating water. The final scenes each span approximately $2000\,\mathrm{km} \times 50\,\mathrm{km}$ in the along and across-track direction.

3.3 Ice hydrometeor particle models

230 At microwave and sub-millimetre wavelengths, scattering by atmospheric ice has a significant effect on measured radiances. Radiative transfer simulations for ICI therefore require assumptions on the optical properties of frozen hydrometeors. Six distinct particle models were implemented for this purpose, where the term particle model refers to a particle habit, an associated particle size distribution, and a range of scaling factors to imitate orientation effects. The six particle models are shown in Table 2. A greater number of particle models are now used in simulations for the new database than used for the preliminary

235 database.

Since scattering calculations of ice hydrometeors are computationally expensive, it is preferable that the single scattering properties are pre-calculated. Additionally, ice hydrometeors are highly variable in shape (O'Shea et al., 2016). Taking these factors into consideration, the chosen habits reflect a sample of the standard habits present in the ARTS single scattering database (Eriksson et al., 2018), which offers single scattering data at microwave and sub-millimetre wavelengths for hydrom-

240 eteors ranging from simple liquid spheres to more advanced ice aggregates.

The three particle models denoted with 'AA' in Table 2 each use a mixture of two distinct habits: one for smaller particles and one for larger particles. To represent large particles, aggregates are used. The smaller particle habits are single crystals that correspond in shape to the aggregate they are paired with. For example, the large plate aggregate habit is complemented with the thick plate crystal. These three particle models are intended to be generic, exhibiting different levels of extinction

245 (Ekelund and Eriksson, 2019). The name used to represent the habit mixture under 'Habit' in Table 2 refers to the name of

the large particle habit, i.e. the ARTS aggregate (AA) habit. The remaining three habits used are an attempt to cover specific hydrometeor categories, e.g. graupel.

Particle size distributions (PSD) are sourced from Field et al. (2007) (represented as F07 and available for both tropics and mid-latitudes) and Delanoë et al. (2014) (represented as D14). The inclusion of D14 is a new addition to the database. A single
250 moment version is used, with the coefficients for the modified gamma taken from Table 4 in Delanoë et al. (2014), under 'All (DARDAR)'. The parameterisation for $N_0^*(T)$, the intercept parameter of the normalised size distribution, follows 'DARDAR' in Table 5 of Delanoë et al. (2014).

The assumption of totally random orientation (TRO) of frozen hydrometeors is frequently made in radiative transfer simulations of microwave observations. In reality, microwave imagers observe polarisation differences, i.e. differences in measured
255 brightness temperatures between vertical (V) and horizontal (H) polarisations (Defer et al., 2014; Gong and Wu, 2017). This is consistent with the presence of oriented ice particles. However, the small simulated polarisation differences that arise when using TRO are far from realistic. Therefore, we implement an approximation of azimuthally random orientation (aARO) according to the scheme developed by Barlakas et al. (2021) and extended by Kaur et al. (2022). In the scheme, orientation effects are mimicked by applying a scaling factor to the extinction values obtained from the TRO assumption. The scaling
260 differs between V and H polarisations. This allows for extinction values resulting from azimuthally random orientation to be approximated. The scaling factor for a given simulation is randomly selected from within a pre-determined range, denoted as the aARO factor in Table 2. Kaur et al. (2022) demonstrated that the best performance at microwave wavelengths (166 GHz) was achieved when using a uniform distribution between 1.0 and 1.4, although the result was somewhat dependent on microphysics assumptions and thus may be adjusted according to the choice of habit. To account for the presence of sub-millimetre
265 wavelengths in ICI and thus potentially larger polarisation differences, the distributions were extended to 1.6 for most particle models.

Particle habit and PSD can both depend strongly on environmental conditions and cloud-type (Yau and Rogers, 1989) and therefore do not occur at the same rate (O'Shea et al., 2016). Additionally, the parameters describing the assumed particle model (including the aARO factor) can have a strong impact on the mapping of IWC from reflectivity inversions (Ekelund
270 and Eriksson, 2019). In light of these considerations, the six particle models are not used with equal frequency within the simulations. Instead, the selection of particle model to use within the simulations in a given scene is a weighted random selection process, where the probability of selecting a particle model follows a non-uniform probability distribution. To assign the selection probabilities, a batch of reflectivity-IWC inversions were performed for each of the particle models. An overall distribution of IWC at an altitude of 8 km was constructed by weighting the IWC distributions from each individual particle
275 model at the same altitude. The weights were chosen according to three criteria. Firstly, that each particle model occurs in a minimum of 10% of the database cases. Secondly, that the particle model 'AA1' occurs in a minimum of 30% of the cases, motivated by its strong performance in previous studies (Ekelund et al., 2020; Pfreundschuh et al., 2020; Geer et al., 2021; Kim et al., 2024). Finally, the weighted IWC distribution should agree with an observed distribution of IWC, so far as the previous two criteria allow. The observed distribution was taken from the CloudSat-based product DARDAR v3.1.

3.4 Radiative transfer

This section outlines the calculation of monochromatic pencil beam brightness temperatures. The weighting of these with sensor response data is discussed in the subsequent section. The radiative transfer calculations were performed using the Atmospheric Radiative Transfer Simulator (ARTS) (Buehler et al., 2018).

The static forward model input includes absorption data. The data are taken from the following sources: Nitrogen (Liebe, 1993), oxygen (Rosenkranz, 1993), water vapour (Mlawer et al., 2012), and ozone (Pickett et al., 1998). The choice of absorption models for oxygen and water vapour followed recommendations of an EUMETSAT study updating and developing RTTOV for sub-millimetre (Fox et al., 2024).

At frequencies below 325 GHz, surface contributions can be non-negligible and the surface emissivity must be considered. Over ocean and land, simulations use the Tool to Estimate Sea-Surface Emissivity from Microwaves to sub-millimetre waves (TESSEM^2) (Prigent et al., 2017) and the Tool to Estimate Land-Surface Emissivities at Microwave Frequencies (TELSEM^2) (Wang et al., 2017). TELSEM^2 provides full emissivity parameterisations up to 85 GHz for all land and sea-ice surfaces, except for new and first-year ice where an extension up to 183 GHz is made. Above these frequencies, constant surface emissivities are generally assumed. However, Harlow and Essery (2012) observed increasing emissivity with frequency for stratified snow and decreasing emissivity with frequency for new snow. Additionally, at higher frequencies, Wang et al. (2017) found some disagreement between retrieved sea-ice emissivities from ISMAR observations and TELSEM^2 sea-ice emissivities. Therefore, we developed an empirically-based probabilistic model for snow and sea-ice surface types. Emissivities at both V and H polarisations are generated from a multivariate Gaussian distribution. The ranges of the distributions were determined using emissivities from Hewison et al. (2002); Harlow (2009); Harlow and Essery (2012). The means, variances, and correlation matrices of the distributions are built on emissivity retrievals performed in Munchak et al. (2020). At frequencies 183 GHz and greater, we use Risse (2021) for reference values. Beyond 243 GHz, constant emissivities are assumed.

It was judged too computationally expensive to perform all calculations with a full scattering solver. Instead, two sets of monochromatic pencil beam calculations were performed. Clear-sky radiances were calculated with nominal humidity or with relative humidity fixed to 50%. These calculations excluded all hydrometeors and were performed for multiple frequencies inside each of ICI's passbands. All-sky simulations were made with ARTS' interface to the DISORT scattering solver, at one frequency per passband. For reference, a clear-sky DISORT run was also made. How these calculations were combined to obtain final values is discussed in the following section.

3.5 Sensor characteristics

Radiances observed by the ICI antenna depend on the frequency response and the angular response of the sensor, and both of these factors must be taken into consideration.

Measured spectral response functions vary within the channel passbands. Therefore, if a constant spectral response function is applied, a bias in brightness temperature may occur. The effect was seen to be strongest for channels 06V (325.15±3.50

GHz) and 09V (448.00±7.20 GHz). To avoid such bias, the full spectral response functions, based on measured ICI data, are now included for all channels.

As briefly discussed in Sect. 3.4, clear-sky calculations are performed for multiple frequencies. The spectral radiances must be obtained on a frequency grid with a resolution that is fine enough to capture variation of the spectral response function over the passband (resolution of 10 MHz). To lessen computational load, simulations are run for a reduced number of frequencies across the sideband and then interpolated onto the finer grid. The number of frequencies per sideband to be simulated was allowed to vary between channels. The number for each channel was chosen such that the resulting uncertainties contribute less than 5% of NEΔT. The number needed to fulfil this criterion largely depends on the contamination of ozone at the given frequency, e.g. the 664 GHz channel requires the highest amount at 25 frequencies per sideband. Due to the high computational cost, all-sky simulations are performed using only the centre frequency of the two sidebands.

After application of the spectral response function, spectral radiances are converted to a brightness temperature for a given channel using the inverse Planck function. The conversion of spectral radiances to brightness temperatures can be well approximated by the linear relationship

$$T_b = \frac{B^{-1}(f_{centre}, R_i) - b_i}{a_i},$$ (1)

where B^{-1} is the inverse Planck function, R_i is the spectral radiance for channel i, and f_{centre} is the nominal centre frequency of the channel. a_i and b_i are band-correction coefficients that are derived pre-launch, based on the characteristics of the channel spectral response function. The coefficients were provided by EUMETSAT.

The incorporation of the full two-dimensional antenna response is necessary in order to avoid beam-filling errors. This is a particularly important consideration for ICI, due to the relatively large footprint size. To perform the antenna weighting, data are integrated over the sensor field of view using

$$T_a = \int_\Omega T_b(\Omega)G(\Omega)d\Omega,$$ (2)

where $G(\Omega)$ is the normalised antenna gain function, provided by EUMETSAT, and Ω is the solid angle. The pattern associated with the 183 GHz channel is applied to all channels, to be consistent with the remapping of data within the operational algorithm. T_a is a simulated antenna temperature that matches a remapped observation on the sub-point of the satellite orbit.

To accurately estimate T_a, the T_b field over the antenna footprint should be relatively dense. However, achieving this with purely simulations would not be computationally feasible. This was solved by using fewer pencil beam simulations, arranged in a pre-determined configuration within a footprint. The simulations are then interpolated onto a finer grid. Within the convex hull of the simulations, a linear interpolation is performed. Outside this region, a nearest neighbour extrapolation is applied.

The configuration of pencil beams will have an impact on the accuracy of antenna weighted data, but a high number of simulations is undesirable. To achieve better efficiency of the ARTS calculations, the three-dimensional scenes were divided into parallel two-dimensional slices oriented in the along-track direction. As a consequence, the arrangement of pencil beams consists of a number of tracks corresponding to the two-dimensional slices. In order to determine a configuration that gives a reasonably low final error, a number of combinations of tracks and spacing along-track were tested. Once simulations were

345   performed for a given configuration, the interpolation onto the finer grid was performed and Equation 2 was applied. To provide
a reference value for the error calculations, T_a was also calculated using simulations performed for each point on the finer grid.

The chosen configuration used simulations spaced approximately 4.4 km apart along the centre track, i.e. the satellite sub-
point. This spacing was successively doubled for each track away from the centre. A total of 38 pencil beam simulations spaced
across 11 tracks are used to represent the spatial variation of brightness temperatures in each final antenna temperature. Overall,

350   the reduction in the number of simulations achieved approximately zero bias and a standard deviation of 0.5 K. This standard
deviation is an average over all cases tested, and is generally low for clear-sky cases but increases with hydrometeor impact.

To compute the T_b field at a specific location, a target position is selected. The position is located along the centre track,
imitating the boresight of the antenna directed towards the target. Targets were selected such that the along-track distance
between the database T_a values is approximately 8.9 km. As a result, there is some overlap in the antenna patterns used to

355   calculate each T_a. Upon selection of a target, the sensor position must be calculated. A challenge then arises when deciding on
which incidence angle to use. While all channels produce a common footprint pattern, their instantaneous footprints at surface
level are not located at the same position due to variation in incidence angle (Table 1 of Eriksson et al. (2020)). To address
this, the same ground-level footprint was assumed for all channels. The position of the sensor was then adjusted based on the
relevant incidence angle, accommodating the variation between channels.

360   The re-gridding of brightness temperatures is performed onto an area bounded by ($\pm 0.7°$, $\pm 0.8°$) in zenith and azimuthal
angle relative to the sensor line of sight, respectively. This is approximately equivalent to twice the full width at half maximum
(FWHM) of the antenna pattern. Each set of clear-sky and all-sky simulations are weighted with the antenna response function
separately.

The measurement vector, to be used in the operational L2 retrieval for ICI, consists of cloud signals for each ICI channel. The

365   cloud signal is obtained by subtracting a clear-sky simulation from the observed antenna temperature. The clear-sky antenna
temperature is simulated using RTTOV, incorporating geophysical data from ECMWF. Since humidity data from ECMWF can
be unreliable, a fixed relative humidity is used within the simulations. Further details can be found in Sect. 3.4.3 of Eriksson
et al. (2020).

The objective for generating a cloud signal is to decrease the contribution from the background. In the case of the database,

370   a cloud signal can be obtained from the difference between an all-sky simulation and an all-sky simulation excluding hydrom-
eteors, i.e. the clear-sky DISORT run introduced in Sect. 3.4. This value can be viewed as a 'true' cloud signal, but there is a
risk that uncertainties arising from non-perfect humidity data are introduced. To adjust for the inclusion of ECMWF data and
be consistent with the measured ICI cloud signal, the quantity ΔT_a is introduced. Within ΔT_a, the two clear-sky runs are also
included:

375    $$\Delta T_a = T_{a,cs} + T_{a,as} - T_{a,as,no-hm} - T_{a,cs,fixed-rh}, \tag{3}$$

where $T_{a,cs}$ refers to a clear-sky antenna temperature, $T_{b,as}$ to all-sky, $T_{a,as,no-hm}$ to a simulation with no hydrometeors
included, and $T_{a,cs,fixed-rh}$ to a simulation with fixed relative humidity.
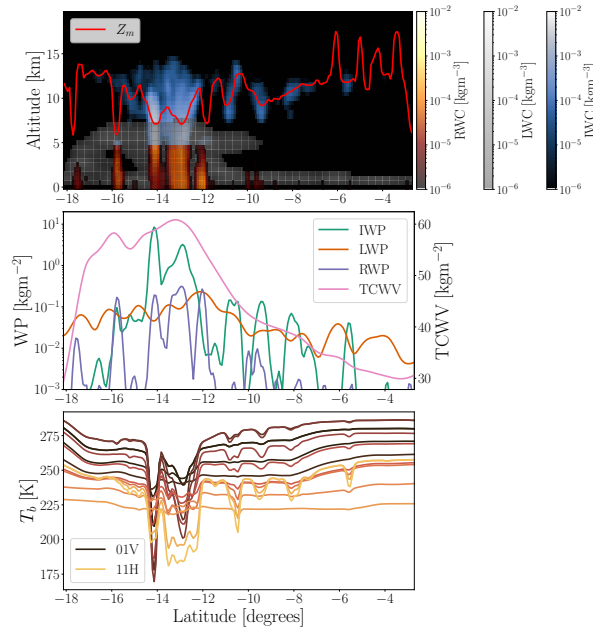
Figure 2. Representation of simulation outputs across the flight direction of CloudSat. IWC, RWC and LWC (top) are not included in the cloud radiation retrieval database to be used operationally, but calculated as an intermediate step. Also shown are Z_m, IWP, LWP, RWP, TCWV, and T_a for all ICI channels, where Table 1 can be referred to for variable definitions. T_a lines are plotted with a gradient colour, where increasing lightness of colour indicates increasing channel ID (see Table 1 in Eriksson et al. (2020)).

Overall, passband properties are captured well in clear-sky simulations. An increasing amount of scattering in simulations (i.e. higher ΔT_a) leads to progressively worse representation of the passband and thus additional simulation uncertainty. How-
380 ever, this uncertainty is small compared to uncertainties arising from other aspects of the simulation, such as hydrometeor assumptions and the neglect of some three-dimensional effects (Barlakas and Eriksson, 2020).

3.6 Output

The database consists of $\sim 9.4 \times 10^6$ cases, where each case is a set of simulated ICI radiances within a remapped footprint, accompanied by aforementioned cloud ice variables. Simulations were performed using 5×10^4 atmospheric scenes. Each
385 database case consists of the quantities given in Table 1. An example of simulation outputs for a given swath is shown in Fig. 2.

4 Retrieval approach

The retrieval of atmospheric quantities is an ill-posed problem in the sense that for a given observation there may be more than one solution describing the underlying atmospheric state, due to the associated uncertainties. Observations uncertainties are

390   generally known only statistically, implying that a retrieval method that retrieves only a single quantity from an observation is unsuitable for the problem at hand. Instead, it is more appropriate that the retrieval is given in the form of a probability density function (PDF) (Rodgers, 2000).

A Bayesian approach is the obvious method if one wishes to perform probabilistic retrievals. Within a Bayesian framework, the solution to the inverse problem is provided in the form of a posterior distribution $p(x|y)$, representing the distribution

395   of possible states conditional on the observation. Since the posterior distribution represents our knowledge of the retrieved quantity after taking into account observations, prior beliefs, and measurement and model uncertainties, it offers a natural way to characterise the uncertainty of the retrieval.

Commonly used inversion methods such as the optimal estimation method (OEM) do adopt a Bayesian framework, albeit a simplified version that produces an approximately Gaussian posterior. However, this becomes unsuitable in the presence of the

400   non-Gaussian statistics and the non-linearity of the forward model that often characterise the retrieval problem (Pfreundschuh et al., 2018).

Bayesian Monte Carlo integration was shown to be successful in the context of ICI retrievals (Eriksson et al., 2020) and will be used in ICI-side of the L2 product at EUMETSAT. However, after the operational algorithm was finalised, quantile regression neural networks (QRNN) began to emerge as an alternative. QRNNs have been shown to perform successful probabilistic

405   retrievals (Pfreundschuh et al., 2018) in the sense that they allow for the estimation of a distribution rather than a single point estimate. Several studies have successfully implemented QRNNs in the retrieval of cloud ice products (Amell et al., 2022). The method also has the potential to perform inversions of an entire swath, since it can make use of the spatial information from observations within the area of interest (Pfreundschuh et al., 2022a). There are other retrieval approaches that allow for uncertainty quantification. One such approach is to use a Bayesian Neural Network (BNN). BNNs have the strength of being

410   able to estimate both aleatoric and epistemic uncertainty, and have been successfully used to retrieve of IWP and D_m (Dong et al., 2023). However, Pfreundschuh (2022) observes that aleatoric uncertainty is often the dominant source of uncertainty in the retrieval problem, which QRNN are suited to handle. Furthermore, BNNs take more time to train and evaluate. Since QRNN has established itself as a faster and more flexible approach to the retrieval problem, it became the obvious choice for our test ICI retrievals.

415   **4.1 Machine learning**

Machine learning problems are typically used to predict an output $\boldsymbol{y}$ from a given input $\boldsymbol{x}$ such that $f(\boldsymbol{x}) = \boldsymbol{y}$. The model f learns to map input to output from a training set $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_i^n$.

A quantile regression neural network (QRNN) is a machine learning approach designed to provide an estimation of the quantiles of the cumulative distribution function of a prediction output value. The τth quantile x_τ of a cumulative distribution

420   function $F(x)$ can be defined as

$$x_\tau = \inf\{x : F(x) \geq \tau\}, \tag{4}$$

where $\tau \in [0,1]$ (Pfreundschuh et al., 2018). A QRNN is essentially an extension of machine learning regression models trained to predict the median. To this end, a QRNN employs a quantile loss function that calculates the loss between predicted and target values for a given quantile level τ. This allows the variability of the prediction to be captured. The quantile loss function is defined as

$$
\mathcal{L}_\tau(x_\tau, x) = \begin{cases} \tau|x - x_\tau|, & x_\tau < x \\ (1-\tau)|x_\tau - x|, & \text{otherwise.} \end{cases}
\tag{5}
$$

The expectation of $\mathcal{L}_\tau(x_\tau, x)$ with respect to x is minimised by x_τ. Therefore, given the training data $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_i^n$, the model f is trained to minimise the expectation of $\mathcal{L}_\tau(f(\boldsymbol{x}), \boldsymbol{y})$. Extending the model to predict multiple independent quantiles allows a discrete approximation of the cumulative distribution function (CDF) $F_{y|\boldsymbol{x}}(y)$ to be made. From a Bayesian perspective, the PDF derived from the CDF can be thought of as a posterior distribution that incorporates the a priori knowledge inherent in the training data. We note that machine learning terminology is used here; $\boldsymbol{x}$ is generally used in machine learning to represent the input to the model, which in this case is a measurement. The same measurement is typically represented as $\boldsymbol{y}$ when used in the context of the forward model and the inverse problem.

A single model is trained to handle all observations in the retrieval database, i.e. all surface-types and latitudinal regions are included. The input vector $\boldsymbol{x}$ includes observations from the 13 ICI channels, in the form of antenna temperatures T_a. The use of ΔT_a as a model input (as to be used in the L2 product) was also checked, and no major difference in retrieval performance was observed between the two variables (Appendix C). Surface temperature, surface pressure and surface classification are also included within the input as ancillary data. The output of the model are 99 quantiles of the CDF $F_{y|\boldsymbol{x}}(y)$ for each of the variables IWP, Z_m and D_m. Details on the configuration of the model are given in Appendix A.

5 Results and discussion

5.1 Simulation of other instruments

The success of the inversion method is highly reliant on the simulation quality of the database. Due to the lack of real ICI observations, assessing whether the simulations are consistent with reality becomes challenging. The approach taken to address this difficulty was to conduct simulations of similar, existing instruments, namely The Global Precipitation Measurement (GPM) Microwave Imager (GMI), The Microwave Airborne Radiometer Scanning System (MARSS) and The International Submillimetre Airborne Radiometer (ISMAR). The simulated radiances were subsequently compared to real observational data. Since actual flight paths were not simulated and the time period considered differs between simulation and observation, the comparisons were made in a statistical sense. Simulation of GMI allows for a comparison of global simulations, whereas simulation of ISMAR and MARSS functions as a regional study with the advantage of covering sub-millimetre wavelengths.

5.1.1 Simulation of GMI

The GMI instrument is a satellite-borne microwave radiometer operating at frequencies ranging from 10 GHz to 183 GHz. We simulated observations from the four high-frequency GMI channels that measure at 166 GHz and 183 GHz. The two 183 GHz channels were chosen due to the benefit of the overlap with the ICI 183 GHz channels. The two 166 GHz channels measure at V and H polarisation and thus offer the opportunity to test the improvements to the simulation framework with regard to polarisation and particle orientation. The simulations used GMI sensor characteristics, i.e. incidence angle, spectral response function and footprint size. To account for GMI instrument error, noise was added to the simulated T_a in the form of a zero-mean Gaussian with standard deviation according to Table 1 in Kaur et al. (2022). Since GMI has only been operational from 2014, simulations are compared to GMI observations from 2020.

Both the one- and two-dimensional distributions of T_a show good agreement between observations and simulations, although a few small discrepancies are present (Fig. 3). In the one-dimensional distributions, the peak at high T_a corresponds to clear-sky cases. There is strong agreement in the peaks. However, the observations reach slightly higher temperatures in the case of the 166 GHz channels. It was found that these cases occurred in desert regions. One possible reason for this difference could be that the emissivities assumed in the simulations are too low. TELSEM land emissivities are derived only for 85 GHz, and it is not known if the extension to 166 GHz remains valid. A second potential reason is that the ERA5 skin temperatures used in the simulations are too low in desert regions.

Good agreement is also seen for most cases with cloud impact (lower T_a). There is some disagreement at the lowest temperatures, where it can be seen that cases of $T_a < 100$ K are present in the observations and not in the simulations. However, these observed cases occur below a probability density of 10^{-6} K^{-1}. Since the simulation distributions are computed with only $\sim 7 \times 10^6$ cases, it is unsurprising that scenarios of $T_a < 100$ K are not well represented in the simulations.

The aim of simulating GMI is not necessarily to achieve perfect agreement with observations, since we do not simulate the same time frame and flight paths that the observations cover. Rather, it is important that the simulations fully cover the variability of the observations. Aside from the region of $T_a < 100$ K and $T_a > 300$ K, this holds true.

When comparing the joint distributions, simulations and observations are in good agreement on both the extent of the correlations and the location of the contour levels. Again, the observed distributions extend to lower temperatures than the simulations, which is expected due to fewer simulations in total. There is a clear difference between the 166V and 166H channel T_a. This implies that the introduction of the approximation of azimuthally oriented particles in the new database produces the required effect in the simulations. To further check the agreement on polarisation effects, the polarisation difference was plotted against the 166V channel T_a, across a range of surface types. The resulting figures (not shown) were very similar to Fig. 4 in Kaur et al. (2022). Likewise, good agreement between the simulations and observations was found, confirming that the simulations capture the variability in polarisation difference in the case of the 166V and 166H channels.
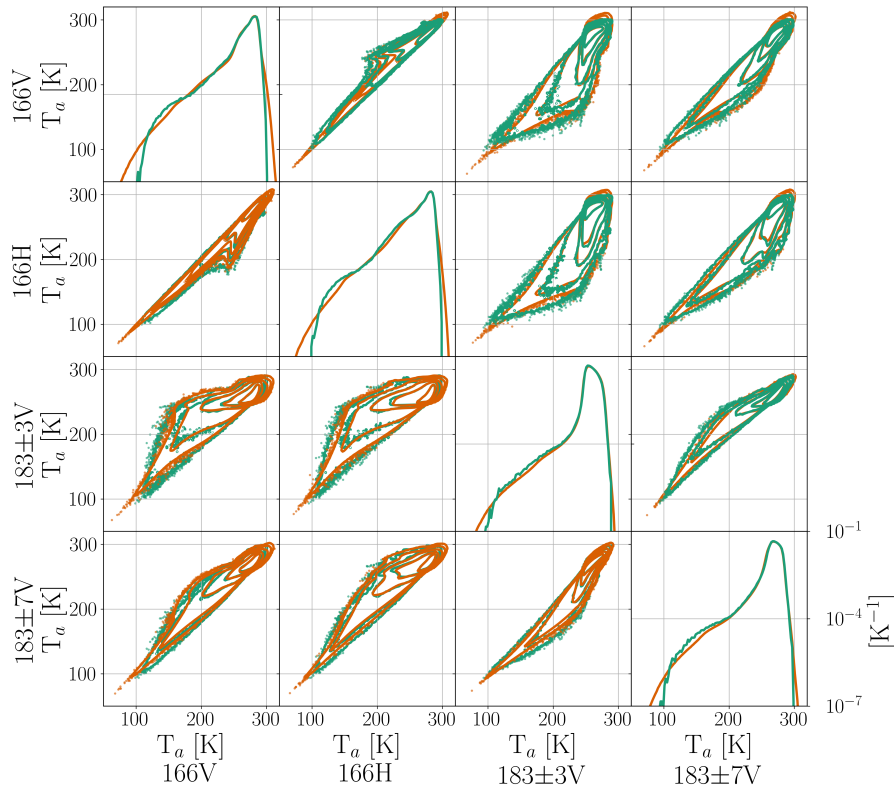
Figure 3. Distributions of simulated (green) and observed (orange) antenna temperatures T_a for four high-frequency GMI channels, simulated using the database generation framework developed for ICI. The plots in the upper right triangle are identical to those on the lower left triangle, but simply have the simulations plotted over the observations, rather than under. Contour lines in the joint distributions (off-diagonal) correspond to $(10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7})$ K^{-2}. Samples that occur in T_a bins with a probability density of less than 10^{-7} K^{-2} are represented by the scatter points in the joint distributions. GMI observations are taken from the year 2020. The simulations and observations both lie within the latitude range of [-60°,60°.]

5.1.2 Simulation of ISMAR and MARSS

ISMAR is a sub-millimetre radiometer that operated on flight campaigns that serve as airborne demonstrators for ICI (Fox et al., 2017). Also on board the aircraft were MARSS radiometers (McGrath and Hewison, 2001), providing measurements at microwave wavelengths. Together, the observations cover the range of ICI channel frequencies. In the case of the 325.15±1.5
485 GHz channel (I-2-V), only observations prior to 2019 were used due to issues with later flights.
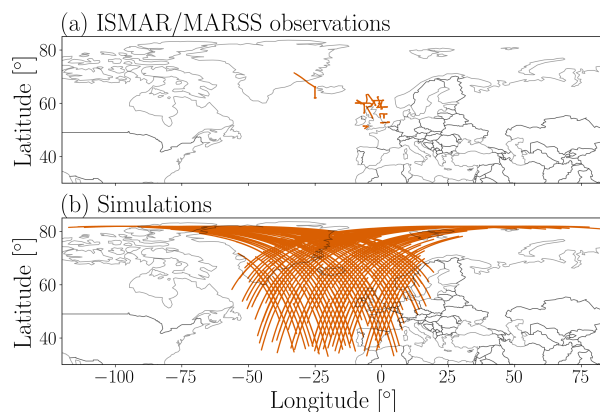
Figure 4. Locations of observations from the ISMAR and MARSS flight campaigns are displayed in panel (a), using cases where flight altitude was greater than 9 km. Locations of the simulated ISMAR and MARSS observations are displayed in panel (b).

Individual flights were not simulated. Instead, simulations were performed within a subset of the three-dimensional scenes used for creating the ICI retrieval database (similarly to the GMI simulations). The location of the simulations and the observations used is shown in Fig. 4. Noise was added to simulations according to Table 3 in Fox et al. (2017).

Similarly to the GMI simulations, the aim of this comparison is to verify that the simulated data spans the measurement space of real observations. Generally, the variability of brightness temperatures present on the flight campaigns is well captured by the simulations (Fig. 5). There are a small number of cases where observations lie outside the simulations. However, it is likely that some observations suffer from higher noise than specified in Fox et al. (2017), possibly explaining the discrepancies. This is most evident in the distributions of channel I-3-V (325.15±3.50 GHz).

Similar features are present in both the simulations and the observations. For example, there are two distinct 'arms' of data visible in many of the plots in Fig. 5. In the case of channels with low surface sensitivity, the two arms correspond to clear-sky and cloudy cases. This can be seen clearly in the joint distribution of M-19-H and I-4-V (183.25±3.00 GHz and 325.15±9.50 GHz, respectively). The clear-sky cases follow the identity line. The arm that diverges from the identity line is a product of scattering due to ice hydrometeors. It extends to increasingly colder T_b as the difference between channel frequencies increases. Such an effect is attributed to the scattering strength increasing with frequency. However, this is not always the case; the M-20-H channel (183.25±7.0 GHz) produces lower temperatures relative to the I-3-V channel (325.15±3.50). This is due to the lower frequency channel having higher atmospheric transmission than the high frequency channel and therefore interaction with a larger fraction of the column of hydrometeor ice.

In channels sensitive to the surface, the arm arises from the surface contribution. The effect is most evident in channels at the same frequency but different polarisations, such as for channels I-1-V and I-1-H (243.20±2.50 GHz), where a significant polarisation difference can be seen. This is more evident in the case of the simulations, where the polarisation difference reaches 50 K. Such high polarisation differences are attributed to dry conditions over the ocean, of which there are significantly more cases present in the global simulations than in the limited region covered by the flight campaigns, as shown in Fig. 4.
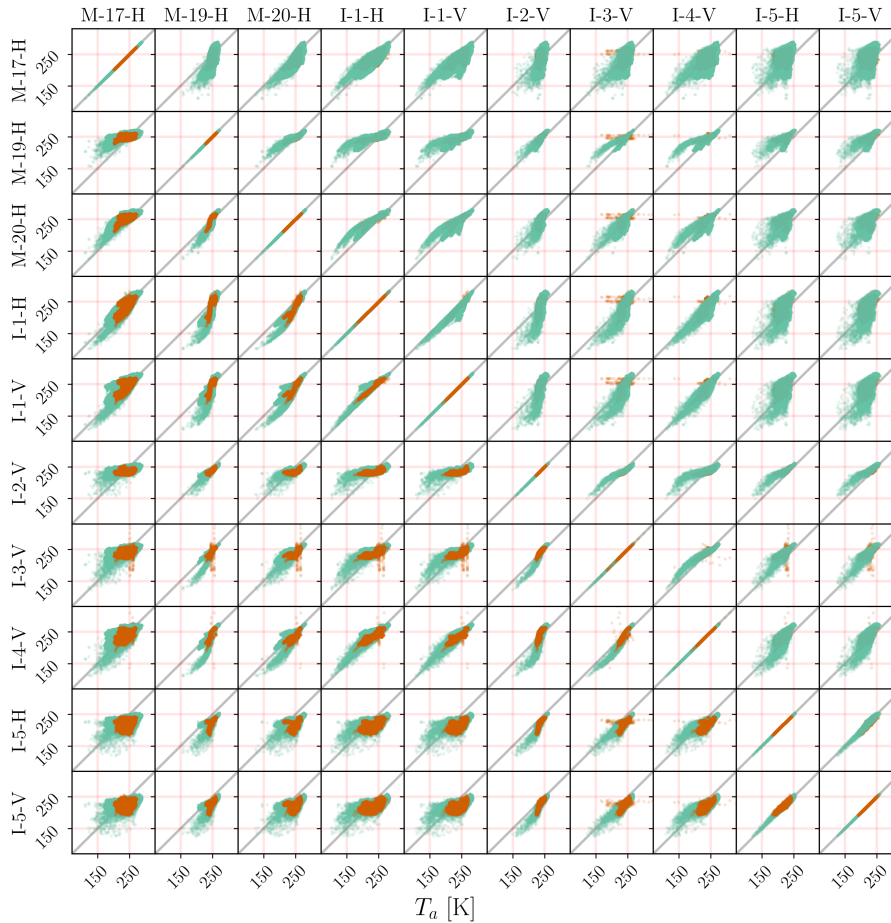
Figure 5. Comparison of observations (orange) from the ISMAR and MARSS flight campaigns and simulations (green) of ISMAR and MARSS using the database generation framework. The plots in the upper right triangle are identical to those on the lower left triangle, but simply have the simulations plotted over the observations, rather than under. The channels are labelled according to the channel IDs given in Table B1.

5.2 ICI Database radiances

One- and two-dimensional distributions of simulated T_a across all ICI channels are displayed in Fig. 6. A relatively large
spread can be observed in the joint distributions between most channels, suggesting that all channels contain some independent information. It can also be noted that in the case of the 243.2 GHz channels (04V and 04H) and the 664 GHz channels (11V and 11H), the two-dimensional distributions show a polarisation difference (the difference between vertically and horizontally polarised channels), which was absent in the preliminary database. The polarisation difference is displayed as a function of the vertically-polarised T_a in Fig. 7. Horizontally polarised channels generally result in colder T_a, leading to predominantly positive polarisation differences. The 243 GHz channel is sensitive to the surface, and therefore polarisation differences due
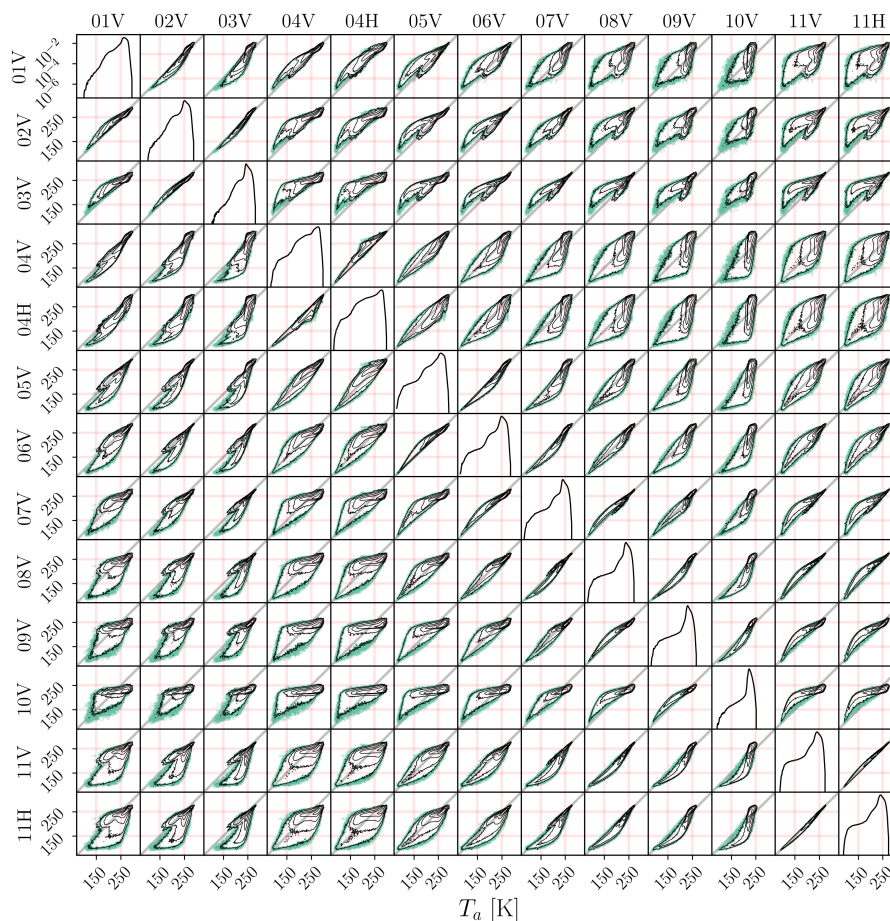
Figure 6. Distributions of simulated brightness temperatures for all ICI channels. Contour lines in the joint distributions (off-diagonal) correspond to $(10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7})$ K^{-2}. Samples that occur in T_a bins with a probability density of less than 10^{-7} K^{-2} are represented by the green scatter points in the joint distributions.

to both surface and cloud impact can be seen in panel (a). Polarisation differences arising from surface interactions (high T_a cases) extend higher than cloud-impacted cases. The features discussed in Sect. 5.1.2, i.e. the two distinct arms of data, are also present in the ICI simulations, confirming that the simulations are capable of emulating real physical behaviour.

5.2.1 Degrees of freedom

520   Retrievals will depend on the sensitivity of ICI observations to atmospheric conditions, above the instrument noise level. To assess the sensitivity, we performed an analysis of the degrees of freedom (DoF) of the simulated ICI radiances. The DoF was calculated as a function of IWP and water vapour across all latitudes for both the new database and the preliminary database.
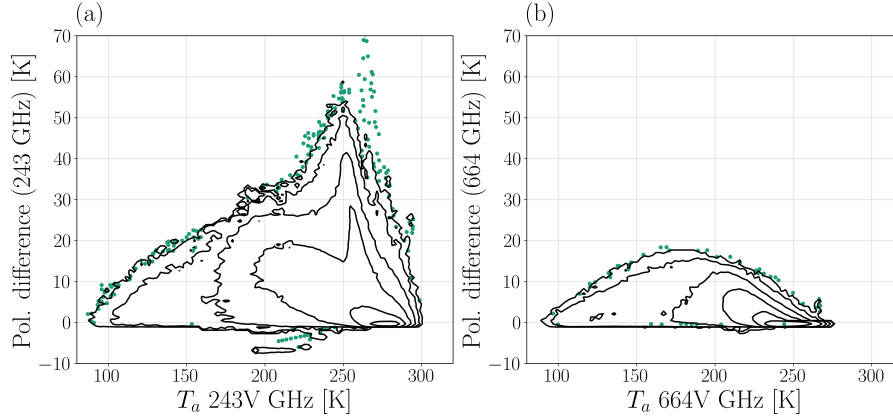
Figure 7. Polarisation differences as a function of T_a at V-pol, plotted as contours of the two-dimensional normalised probability density function. The distribution at 243 GHz is shown in panel (a) and the distribution at 664 GHz is shown in panel (b). Contour lines correspond to $(10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7})$ K^{-2}. Samples that occur in bins with a probability density of less than 10^{-7} K^{-2} are represented by the green scatter points.

The DoF quantifies the number of independent variables free to vary, and may be considered to be a measure of the information content of a measurement in the presence of noise.

525   The DoF is essentially a comparison between the variance of the measurements and the variance of the measurement noise, conducted in eigenspace. To achieve this, an eigendecomposition of the covariance matrix $\mathbf{S_y}$ of noise-free simulated T_a is performed:

$$\mathbf{S_y} = \mathbf{E\Lambda E}^T, \tag{6}$$

where $\mathbf{E}$ is the matrix whose columns contain the eigenvectors of $\mathbf{S_y}$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues.

530   The diagonal matrix $\mathbf{S}_\epsilon$ is constructed to hold NEΔT^2 in its diagonal elements, and transformed to eigenvalue space:

$$\mathbf{S_\Lambda} = \mathbf{E S}_\epsilon \mathbf{E}^T. \tag{7}$$

The DoF can then be calculated as the number of eigenvalues that are larger than the corresponding diagonal element of $\mathbf{S_\Lambda}$. In the context of ICI, it can be interpreted as the effective number of channels.

    Panel (a) of Fig. 8 presents the DoF of the simulated ICI observations as a function of IWP and water vapour. The overall
535 trend in DoF is consistent with Figure 8 in Eriksson et al. (2020), although the DoF is now calculated across all latitudes, rather than only the tropics. An increase in either water vapour or IWP corresponds to an increase in DoF, although the change in DoF is much larger over the range of IWP. The general conclusion drawn from this result is that ICI is sensitive to humidity, but demonstrates strong sensitivity to ice hydrometeors.

    We applied the same DoF analysis to both the new database and the preliminary database, and the resulting difference in
540 DoF is presented in panel (b) of Fig. 8. Overall, there is an increase in DoF; a positive difference in DoF is present for all but

very humid, clear-sky conditions, and there never occurs a decrease in DoF. The DoF changes depending on both water vapour and IWP conditions, ranging from 3 (clear, humid conditions) to 10 (very cloudy, humid conditions). The change relative to the preliminary database ranges from 0 to 3. In order to understand this variation in DoF, the regions of Fig. 8 must be considered separately.

545   In the scenario of high water vapour and low IWP, a DoF of 3 to 4 is found. This is consistent with the three channels that surround each of the water vapour molecule transitions covered. A total of nine ICI channels cover three transitions, leading to redundancy between these channels and thus a relatively low DoF. The majority of this region sees no change in DoF relative to the preliminary database, although a small increase of 1 DoF can be seen for the highest water vapour cases.

    Under conditions of low water vapour and low IWP, the DoF increases to 5 and 6 due to increased sensitivity to lower levels
550 of the troposphere and the surface. This corresponds to an increase of 1 to 2 DoF from the preliminary database. This is likely due to the inclusion of V and H polarisation for the 243 GHz channel, allowing for surface effects to be better captured.

    For IWP ≥ 40g m^{-2} (slightly above ICI's sensitivity threshold), the DoF does not go below 6. A steady increase in DoF occurs with increasing IWP. When comparing to the preliminary database in the region of IWP above the sensitivity threshold, an increase of 1 to 2 DoF is achieved throughout.

555   The maximum DoF found is 10, occurring in the region of very high IWP and water vapour. This corresponds to an increase of 3 DoF relative to the preliminary database. Although a DoF of 13 would be consistent with the number of channels, such a high value is not necessarily expected due to correlation between the channels (see Fig. 6) and the higher noise present for some channels decreasing their contribution. However, as discussed previously, a relatively large spread can be seen in the joint distributions between all channels, suggesting that every channel contributes to the DoF.

560   To investigate the contribution of channels to the DoF, we recalculated the DoF across four datasets, each excluding channels centred on a single frequency: 243 GHz, 325 GHz, 448 GHz, or 664 GHz. When excluding the surface-sensitive 243 GHz channels, a significant decrease in DoF occurred for conditions of low water vapour and low IWP, i.e. conditions at which the surface is somewhat visible. The 325 GHz channels and the 448 GHz channels cover water vapour transitions. As expected, removing either of these sets of channels led to a decrease in DoF for all low IWP cases across the entire range of IWP.
565 Exclusion of the 325 GHz channels also lowered DoF for low water vapour and high IWP values, whereas exclusion of the 448 GHz channels lead to a more prominent decrease in the region of high water vapour and low IWP. This suggests that, although both sets of channels cover water vapour transitions, neither set is redundant and both contribute to ICI's overall sensitivity. Finally, removal of the ice hydrometeor-sensitive 664 GHz channel led to a decrease of up to 2 DoF for all high IWP cases. In fact, all channels led to a small decrease of DoF for high IWP values, implying that every ICI channel can play a role in the
570 retrievals.

    To assess the impact of the introduction of polarised channels in the database, we conducted an additional analysis of the DoF on a 'polarisation-free' version of the new database. Namely, we replaced the individual observations from V and H polarised channels at the same frequency with their mean value. A decrease of at least 1 DoF was observed over the entire region of IWP ≥ 40g m^{-2}, with a decrease of 2 in the high-IWP and high-WV region. This leads to a maximum DoF of 8.
575 Despite the decrease in DoF in this region, there still remains an overall positive difference in DoF between the new database
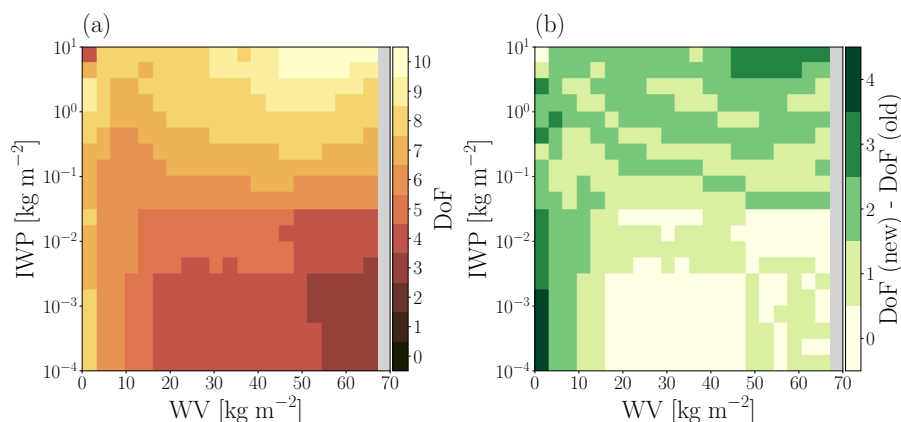
Figure 8. The estimated degree of freedom (DoF) of the simulated observations within the ICI retrieval database is presented in panel (a). DoFs are given across total column water vapour (WV) and IWP. Shown in panel (b) is the difference in DoF achieved with the new retrieval database and the preliminary database found in Eriksson et al. (2020). Cases at all latitudes are included.

and the preliminary database. In the region of low-IWP and low-WV (conditions under which the surface can be seen), a drop of 1 to 2 DoF occurred, resulting in little difference relative to the preliminary database.

To check the impact of now assuming 75% NEΔT, another analysis was performed using the original noise values of NEΔT. The maximum DoF achieved was 8, with an overall decrease in DoF seen across the entire region of high-IWP. Small decreases were seen elsewhere, but these were not major.

Although the analyses performed were relatively simple, they do allow us to infer that both the introduction of polarisation and the reduction in noise have improved the DoF of the database. Additionally, increases to the DoF relative to the preliminary database still occurred when attempting to remove the influence of these factors. We conclude that other developments to the database (see Sect. 3.1) also play a role in the increases of DoF, although further analysis is required to quantify this further.

5.3 Ice water path retrievals

A QRNN was trained on a subset of the database and subsequently used to retrieve IWP, D_m, and Z_m from observations in a separate subset. The success of the retrievals was evaluated through a comparison of retrieved cases to true cases. Evaluating a probabilistic estimate against a single true value can be challenging unless done on a case-by-case basis. To compare a set of retrieved distributions with the 'truth', a point estimate must be selected from each distribution. There are several options for choosing a point estimate. One simple option is to calculate the mean of the distribution. To do so, a CDF is constructed by interpolating between the retrieved quantiles and extrapolating to 0 and 1 outside the minimum and maximum quantiles, respectively. The first moment of the predicted PDF provides the mean. An alternative choice of point estimate is to take a single random sample from a posterior distribution, where the posterior distribution is computed through an interpolation of the inverse CDF. By definition, a mean does not contain information on the values present in the tails of a distribution. In

595   contrast, there is some probability that a random sample from the posterior will fall within this region. Therefore, a large set of point estimates obtained by randomly sampling the distributions is expected to capture more extreme cases.

When comparing retrievals directly to the database values, we use the mean as the point estimate. The mean, median, 16% quantile, and 84% quantile of all retrieved means are subsequently calculated as a function of the database values, primarily to provide a visual comparison. The remaining statistics, such as bias and correlation coefficient, are computed on the complete

600   set of posterior means.

Retrieval performance for IWP is presented in panel (a) of Fig. 9. In the case of IWP retrievals, the correlation coefficient is found to be $r = 0.87$. The overall bias for retrievals of IWP > 1 g m^{-2} is -4 g m^{-2}. We note that the bias can be a misleading statistic due to its capacity to be influenced by the density of samples in a particular region, or if the data ranges across orders of magnitude. Alternatively, Fig.9 offers a qualitative, and perhaps more intuitive, representation of the bias. The median follows

605   the true values closely for 10g m$^{-2} <$ IWP < 1.7 kg m^{-2}. Retrievals at IWP > 1.7 kg m^{-2} display a negative bias, likely due to fewer of these cases present in the training data. Values in this region are the main contributors to the negative bias. The overall bias is therefore small relative to these influential high-IWP values. IWP data less than 1g m^{-2} are not represented in the figure, but have an overall positive bias of 2g m^{-2}. A positive bias appears in the median at IWP < 10g m^{-2} and in the mean at IWP < 15g m^{-2}. The increasing uncertainty and bias of the retrievals in this region can be attributed to the fact that

610   observations become progressively less sensitive to smaller amounts of ice mass, due to the low cloud-impact of these cases. The effect can be seen in the decreasing gradient of the mean and median with decreasing IWP. The overall bias across all IWP retrievals is 0.6g m^{-2}, indicating that there is a trade-off between the negative and positive bias present for low and high IWP, respectively.

Statistical distributions were made for both retrieved cases and database cases, and subsequently compared to DARDAR

615   (Fig. 10). The comparison is motivated by the fact that successful retrievals are meaningful only if the true cases are themselves realistic. DARDAR IWP was obtained by integrating column-wise all IWC data available in the DARDAR v3.1 product for the year 2010. Both of the aforementioned retrieval point estimates were used and compared to the database simulations and to DARDAR. Good agreement is seen for all data in both the overall distribution and the zonal mean. Since the retrieval distributions match the database, this is an indication that the retrievals are successful in a statistical sense. The model does

620   not appear to be biased within the region of 10^{-3} kg m$^{-2} \leq$ IWP $\leq 10^2$ kg m^{-2}. This is supported by the model statistics computed for this region (see Fig. 9).

The database simulations and, by extension, the retrievals achieve a higher probability density of high-IWP values than in DARDAR, which is visible in panel (a) of Fig. 10. The DARDAR distribution drops sharply for IWP > 10 kg m^{-2}, whereas the other distributions continue to higher IWP. However, we do not aim to simply reproduce the variability of DARDAR data.

625   In fact, Cazenave et al. (2019) found that changes to the DARDAR-CLOUD product between versions v2 and v3 led to a 24% reduction in IWP on average. It is not automatically true that v3 is 'better' than v2. This emphasises the fact that DARDAR does not necessarily represent the truth, and therefore achieving a perfect match with DARDAR is not essential. Additionally, simulations suggest the possibility of up to 50 kgm^{-2} in the presence of tropical deep convection (Bolot et al., 2023). It is therefore possible that DARDAR fails to identify the highest IWP cases. Besides, a greater number of high-IWP cases in the
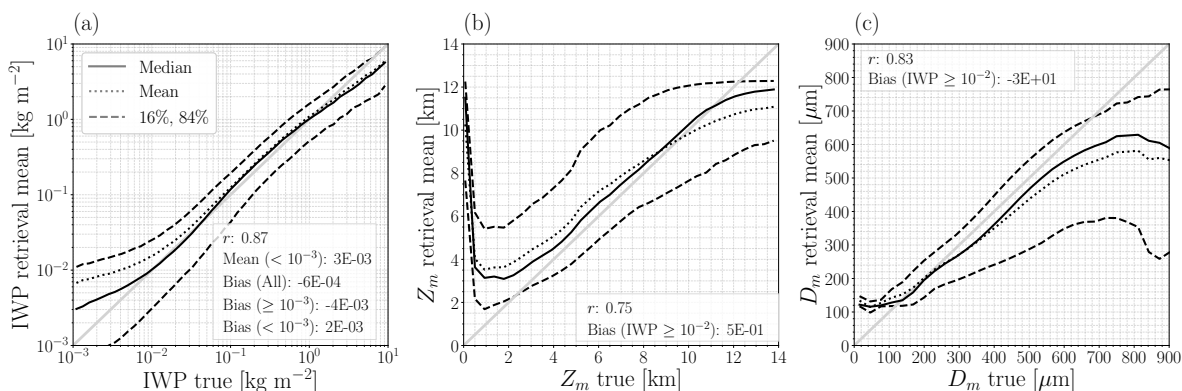
Figure 9. Retrieval performance for IWP, Z_m, and D_m, shown in panels (a), (b), and (c), respectively. The point estimate of the retrieval is taken as the mean of the retrieved posterior distribution. The mean, median, 16th quantile and 84th quantile of the retrieved mean are plotted as a function of the database values (labelled as 'true'). Summary statistics are shown, where r indicates the correlation coefficient. Bias is given in units of kg m^{-2}, km, and μm for IWP, Z_m and D_m, respectively. Cases of IWP $< 10^{-3}$ kg m^{-2} are not represented graphically, but are included in the relevant summary statistics.

630 database is not seen as a cause for concern since the ICI retrievals will benefit from more high-IWP cases a priori; such cases are rare and an increase will lead to a better-trained model in this region. In the case of the zonal mean, all distributions agree relatively well. Retrievals around 50°N $\pm 10°$ are somewhat low, but this could be due to poorer retrieval performance for mid-latitudes (discussed further in Section 5.5).

  The global distribution of IWP in the database, retrieved IWP, and DARDAR is shown in Fig. 11. The same regional features
635 can be seen in all cases, e.g. the intertropical convergence zone (ITCZ). The datasets also agree on regions containing fewer ice clouds, such as over the Sahara and Arabian deserts and the stratocumulus regions over subtropical ocean. The retrievals appear noisier, but this is simply attributed to the fact there are fewer retrieval cases than present in the entire database and, to an even greater extent, DARDAR.

5.4 Z_m and D_m retrievals

640 Retrievals of mean mass altitude Z_m have an overall bias of 0.5 km (for cases with IWP $> 10^{-2}$ kg m^{-2}) and a correlation coefficient of $r = 0.75$ (panel (b) of Fig. 9). The retrievals correspond well with true (database) values for Z_m between approximately 2 km and 12 km, which is similar behaviour to retrievals performed with the preliminary database (Eriksson et al., 2020). Retrievals of Z_m perform poorer outside this range, though this is as expected.

  Retrievals of D_m are presented in panel (c) of Fig. 9. Good agreement is seen for true (database) values below 600 μm,
645 although a slight underestimation is present within the range of 250 μm $< D_m <$ 600 μm. All retrievals corresponding to cases with IWP $> 10^{-2}$ kg m^{-2} contribute to an overall negative bias of -30 μm. The overall correlation coefficient is $r = 0.83$. Retrievals of $D_m \geq 600$ μm display a high negative bias, and appear to be increasingly less sensitive to the true value as D_m
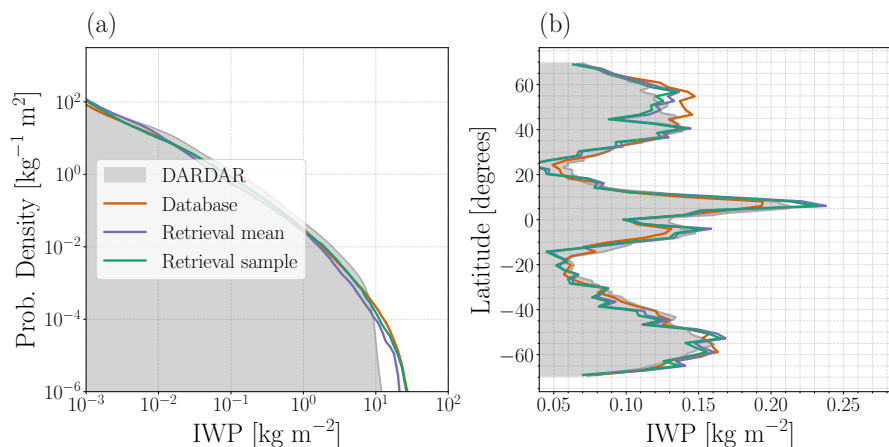
Figure 10. Distribution of IWP cases in the retrieval database and retrieved IWP, shown as an overall distribution and as the zonal mean. The retrieval distributions are calculated taking either the posterior mean as a point estimate or by taking a random sample of the posterior as a point estimate. Also shown is the distribution of IWP calculated as the vertical integral of IWC in the DARDAR product for the year 2010.
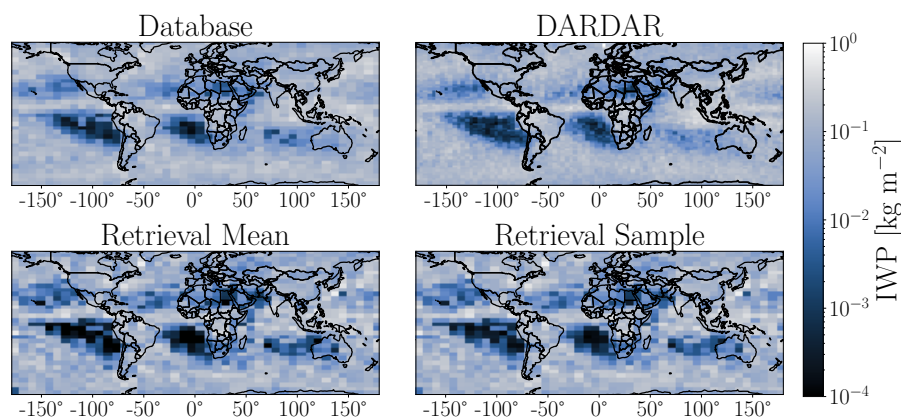
Figure 11. The global distribution of IWP cases in the database, retrieved IWP, and IWP calculated from DARDAR from the year 2010. The plot titled 'Retrieval mean' uses the mean of the retrieved posterior distribution as a point estimate. The plot titled 'Retrieval sample' takes a random sample from the posterior distribution as a point estimate.

increases. This is partially due to the fact that higher-D_m values are associated with lower altitudes, at which less information is available from ICI radiances. However, the decrease in sensitivity can also be a saturation effect in terms of ICI wavelengths.
650 Adding the longer-wavelength MWI observations could improve D_m retrievals in this region. At $D_m < 140\,\mu m$, the retrievals display a significant positive bias, with even the 16th percentile of the retrieved mean falling well above the identity line. The retrievals appear to be unsensitive to the true values within this region.

Overall distributions and zonal means were computed for both database cases and retrieval cases. Results are presented in
Fig. 12. In the case of the zonal means, each case is weighted with its corresponding IWP. Although not directly available in
655 the DARDAR product, Z_m and D_m can be calculated by applying Equations 4 and 5 in Eriksson et al. (2020) to variables that
are available. Z_m and D_m data corresponding to IWP $< 10^{-2}$ kg m^{-2} were excluded from the distribution calculations, as this
data falls below the sensitivity threshold of ICI.

In panel (a), it can be seen that retrievals of Z_m agree statistically with the database (true values) for $Z_m < 12$ km, i.e. for
all clouds of concern for ICI. As previously discussed, taking the mean as a retrieval point estimate can lead to a failure to
660 capture extreme values in a distribution, as can be seen for $Z_m > 12$ km in Fig. 12. However, sampling the retrieved posterior
distribution does produce a distribution that includes higher Z_m cases, with such cases appearing with a probability density
approximately matching that of DARDAR. This acts as a good example of the benefits of retrieving quantiles, rather than a
single estimate.

In the case of the Z_m zonal means in panel (b), database cases, both retrieval estimates, and DARDAR all agree well. This is
665 expected, since Z_m is well constrained by the CloudSat reflectivities used to construct the atmospheric scenes. There are some
small discrepancies when retrieving Z_m at very high and low latitudes, but this may be due to difficulties when distinguishing
between the surface and low-altitude clouds.

The overall distributions of D_m are shown in panel (c). They are similar in shape and agree somewhat well, but tend to
diverge for $D_m > 700$ μm. This was already evident in Fig. 9 when taking the retrieval mean as a point estimate. However, even
670 when sampling the retrieved posterior distributions, we fail to capture the extreme values present in the database, indicating
that the model struggles to retrieve larger particles in general. The difficulty in retrieving high D_m is also present in the zonal
mean, leading to an overall underestimation.

There is one major difference between our simulations and DARDAR: our simulations contain more higher D_m cases
than DARDAR does. This difference is particularly evident in the zonal mean in panel (d). Fortunately, more high D_m cases
675 probably improves the retrieval model in this region. Furthermore, we again stress that DARDAR is not 'truth', but a retrieval
product itself. The discrepancies occur due to differences in particle model assumptions, and serve to demonstrate that we have
not tuned our setup to replicate DARDAR. Overall, our D_m retrievals serve to demonstrate that ICI has sensitivity to D_m. This
result is supported by Pfreundschuh et al. (2020), which demonstrated that sub-millimetre observations can better constrain
microphysical properties.

5.5 Retrieval sensitivity

680

There are multiple factors that may affect retrieval performance. Some effects are related to the sensitivity of the radiances to
certain conditions, such as surface type, climatic region, and the optical properties of the ice hydrometeors. For example, it is
difficult to isolate surface effects from cloud observations, and therefore surface types associated with higher uncertainties are
expected to result in poorer retrievals. Additionally, the latitudinal region will also affect the sensitivity of observations to ice
685 mass, due to varying cloud base altitude. Alternatively, retrieval performance can be impacted by the amount of thermal noise
present in the observations. This section is an investigation into the sensitivity of retrieval performance to the aforementioned
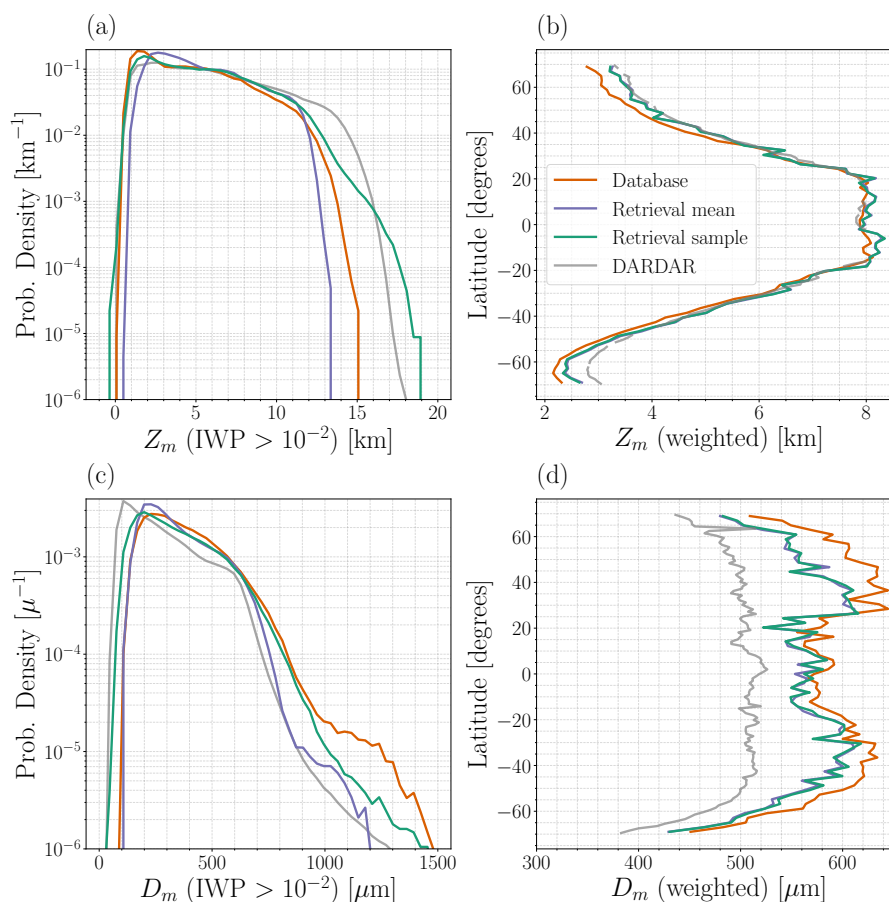
Figure 12. Distributions of Z_m and D_m, computed from database cases, retrieved means, and random samples from the retrieved posterior. Probability density functions are shown for Z_m and D_m in panels (a) and (c), respectively. Zonal means of Z_m and D_m are presented in panels (b) and (d), respectively. The distributions are calculated using only cases corresponding to IWP > 10^{-2} kg m^{-2}, i.e. above the sensitivity threshold for IWP. In the zonal means, Z_m and D_m are also weighted with IWP. Also shown are reference distributions for Z_m and D_m calculated using variables available within the DARDAR product for the year 2010.

factors. Retrieval performance is evaluated for various latitudinal region and surface type, and across the six particle models used in the simulations. Finally, a simplified assessment of the effect of thermal noise on model training is presented.

Retrieval performance is seen to vary according to latitudinal region. Panel (a) of Fig. 13 demonstrates this accordingly. Retrievals at tropical latitudes perform better than at mid-latitudes, in the sense that retrievals and mid-latitudes demonstrate a wider spread between the 16% and 84% quantiles of the retrieval mean. Furthermore, the median line follows the truth closely down to as low as 1g m^{-2}. The finding that retrieval performance improves at tropical latitudes corresponds with results in Eriksson et al. (2020). The decreased performance at mid-latitudes can be attributed to ice clouds lying closer to the surface. To see such clouds, observations must be sensitive to the entire ice column, which is not the case for all channels. However,
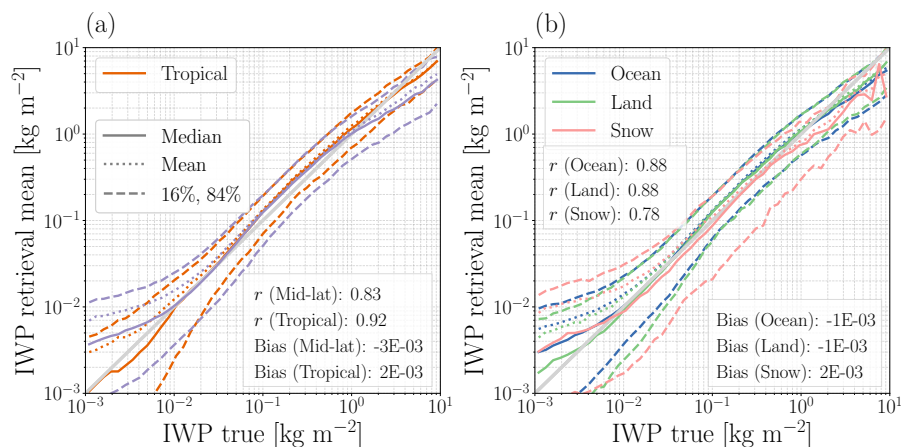
Figure 13. Retrieval performance for IWP, separated by latitudinal region in panel (a) and by surface type in panel (b). The mean of the retrieved posterior is taken as the point estimate. Tropical refers to the region bounded by the Tropic of Cancer and the Tropic of Capricorn, i.e. 23.44°N and 23.44°S. Mid-latitude refers to the region bounded by the Arctic circle and the Antarctic circle, i.e. 66°N and 66°S, excluding the tropical region.

695 channels that are sensitive to the entire column will also produce observations affected by the surface. Surface emissivity effects are still accompanied by significant uncertainties, leading to increasing retrieval uncertainty in the given regions.

The effect of surface type on retrieval performance was also investigated (panel (b) of Fig. 13). Retrieval performance is shown for surface types of ocean, land, and snow. The surface emissivities of snow-covered surfaces remain poorly modelled due to significant variability with frequency. This uncertainty in surface-emissivity, and the fact that this surface type is more

700 common at higher latitudes, will contribute to a poorer retrieval performance over snow-covered surfaces. At IWP $> 10\,\mathrm{g\,m^{-2}}$, retrievals of cases over snow-covered surfaces demonstrate an overall negative bias, and a larger spread between the 16% and 84% quantiles than seen for other surface types. The negative bias is more evident with increasing IWP.

It is difficult to isolate the effect of surface-type when some surface types are more common in certain climatic regions, e.g. snow at high-latitudes. However, retrievals appear to perform better in tropical regions than at mid-latitudes, despite a less

705 pronounced difference in surface-types between these regions. Due to this, and the physical explanations discussed previously, it is likely that climatic region and surface type independently have an impact on retrieval performance.

The bulk optical properties of the ice particle models implemented in the simulations will have varying effects on the observed radiances. The exact relationship is complex, but the effect of scattering plays a major role. An ice particle with a higher degree of extinction will tend to scatter more radiation away from the sensor's line of sight, thus causing a greater

710 impact on brightness temperatures. For example, Geer et al. (2021) found that the level of extinction exhibited by ARTS particle models approximately correlates to the amount of brightness temperature depression. The dense, rimed hydrometeors used in our study (AA3 and graupel) produce the most scattering and thus have the greatest impact on brightness temperature. Conversely, the low density particle (AA2 and snow) have a weaker effect on brightness temperature.
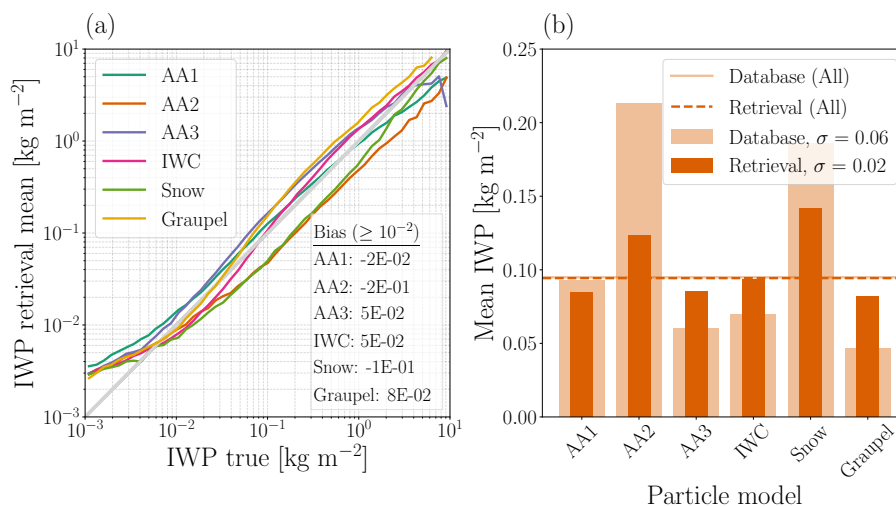
Figure 14. Panel (a) displays the retrieval performance for IWP, separated by the particle model used in the simulated case. The mean of the retrieved posterior is taken as the point estimate, and the solid lines correspond to the median of the retrieved means. Bias is computed for all retrievals corresponding to a true IWP $\geq 10^{-2}$ kg m^{-2}. In panel (b), the bars shown the mean IWP for a given particle model. Statistics were computed across the test set for both simulated (database) cases and retrieved cases. The solid horizontal line represents the mean IWP for all simulated cases in the test set, and the dashed line represents the mean retrieved IWP for the same cases. The standard deviation σ corresponds to the spread of mean IWP between the particle models, and is in units of kg m^{-2}.

In the simulations conducted for this study, clouds are assumed to contain only one type of particle. For a cloud with a given
715   IWC, a highly scattering ice particle will cause a greater depression in T_a. Reciprocally, when simulating a cloud composed of highly scattering ice particles, less IWC is needed to replicate the same T_a when compared to a cloud consisting of particles with weaker scattering.

In an ideal scenario, the retrieval model has the capability to infer the particle model used in a simulation and adjust its IWP prediction accordingly. Alternatively, our model may assume a 'mean particle model' that exhibits the average level of
720   extinction observed across all six ice particle models. If this is the case, then the accuracy of the retrievals can be affected. For instance, if a dense, highly-scattering particle was used in a simulation, the model may overestimate IWP in an attempt to align its prediction with the simulated T_a. The converse will also be true; If a particle with lower relative extinction was used in a simulation, the model may underestimate IWP.

To investigate the sensitivity of the retrievals to the choice of particle model, we computed retrieval statistics for each
725   particle models individually (panel (a) of Fig. 14). Above the sensitivity threshold of IWP = 10^{-2} kg m^{-2}, there is a trend of overestimation in the case of AA3 and graupel. Likewise, IWP retrievals are generally underestimated in the case of AA2 and snow. Based on these findings, we conclude that the retrievals show sensitivity to the particle model used in the simulations. This suggests that the model generally assumes some average level of extinction. In panel (b) of Fig. 14, we present the mean IWP associated with each particle model for the same set of data, for both simulated and retrieved cases. The spread of bar heights

730   indicates variability between the particle models within the simulated (pre-retrieval) data ($\sigma_{\text{pre}} = 0.06 \, \text{kg m}^{-2}$). This variability decreases post-retrieval ($\sigma_{\text{post}} = 0.02 \, \text{kg m}^{-2}$). However, the variability does not disappear completely. In conclusion, it may still be possible that the model can distinguish between particle models, but cannot do so with total reliability. An in-depth analysis would be required for further insights. An interesting avenue for such an analysis would involve attempting to retrieve the particle model itself.

735   Noise is incorporated into the retrievals at two stages; Noise was added to the observations during model training, and to the observations in the test dataset prior to inverting them. To investigate the impact of NEΔT, two tests were made that alter the inclusion of noise. In the first test, the original model, trained with noise, was used to invert the cases in a noise-free test dataset. In the second test, the model was re-trained without noise added to the training observations. Next, noise-free observations were inverted. The second test is equivalent to assuming NEΔT$= 0$, i.e. an ideal but unrealistic observation with

740   no thermal noise. Only a very small increase of r was seen in the first case, with $r = 0.88$ for IWP retrievals. In the second case, a decrease to $r = 0.79$ was observed for IWP retrievals. This is likely due to the model overfitting to the data in the total absence of noise. In both cases, only very minor visual differences were seen in the equivalents of Fig. 9. We conclude that the current values of NEΔT have little impact on the training of our retrieval model. This implies that the retrieval performance discussed in this study can be attributed to the quality of the database simulations and to the sensitivity of ICI, rather than the

745   influence of noise.

5.6   Limitations

Although efforts were made to produce the most realistic ICI simulations available, there remain limitations. Firstly, although a range of particle models are used, melting ice particles are not considered. Single scattering data for mixed phase particles is not currently available for sub-millimetre frequencies, and calculating such properties would be computationally costly

750   (Kanngiesser and Eriksson, 2022). A second limitation arises from the simplification of assuming the same particle model across an entire column or footprint. This is an unphysical assumption (Kim et al., 2024) that likely affects the realism of the simulations. Another limitation relates to the challenge of representing three-dimensional variability in the simulations. To this end, the Barker method was applied. However, it is possible that this method underestimates the true variability.

It is difficult to identify all weaknesses in the absence of real, global ICI data. After the launch of ICI, there will be the

755   possibility to compare to real observations. A comparison of observations and simulations at 243 GHz will allow a better assessment of the surface emissivity parameterisation. On the other hand, a clear-sky comparison using channels with lower surface sensitivity will provide insight into the gas absorption models. It will be more complex to identify limitations for cloud-impacted cases, but an investigation of polarisation differences could provide insight into the success of the aARO scheme. It is noted that, in only ten days, ICI will generate the same number of observations as in the retrieval database. Statistically, there

760   will be observations that lie outside the variability of the simulations. However, if it is found that the simulations clearly do not cover the variability of real ICI observations, investigation will be required.

6   Summary and conclusions

In this study, we presented the state-of-the-art retrieval database developed for use within the operational ICI retrieval algorithm. The database generation strategy was based upon Eriksson et al. (2020), but multiple improvements were made. Among the most significant extensions were the inclusion of the full channel responses and representation of the full two-dimensional antenna pattern. Additionally, the treatment of ice hydrometeors now incorporates an approximate treatment of particle orientation, allowing for polarisation effects to be captured. Even taking into account the approaches taken to improve calculation efficiency, the simulations performed within this study are, to our knowledge, the most detailed performed to date.

In the first part of the study, we performed an analysis of simulated antenna temperatures. The degrees of freedom of the antenna temperatures were computed and results were compared to the same method applied to the preliminary database of Eriksson et al. (2020). An overall increase in the degrees of freedom was seen, indicating that simulations are more realistic and that improvements made to the database have increased the information content of the simulated radiances.

We also simulated radiances as observed by existing instruments, in order to validate the simulation methods while bypassing the need for real ICI observations. The database generation framework was used to simulate antenna temperatures as would be observed by sensors on the flight campaigns ISMAR and MARSS. This allowed for a comparison of both microwave and sub-millimetre observations. The simulations were validated against real observations and were found to cover the entire measurement space. Likewise, we simulated four high-frequency microwave GMI channels. The simulated radiances overall followed the same statistical distribution as real GMI observations, and covered the measurement space in most cases. Some differences were observed, but occurred only for more extreme brightness temperatures, and can partially be explained by too few simulated cases.

In the second part of the study, we characterised the expected retrieval performance associated with using the database within an inversion scheme. A probabilistic machine learning model was trained on the database simulations. The model then performed inversions of simulated antenna temperatures to retrieve ice water path, mean mass height, and mean mass diameter. The results indicate a sensitivity to IWP between 10g m^{-2} and 10 kg m^{-2}. Additionally, we show that both simulated and retrieved IWP and Z_m data are statistically consistent with the DARDAR product. Simulated and retrieved D_m cases slightly differ from DARDAR in their distributions, demonstrating that our retrievals are not simply a replica of another product.

In this study, we decided to retrieve only the variables that will be offered in the EUMETSAT L2 product. However, it is possible that the radiances contain enough information to retrieve other variables. For example, due to the high number of ICI channels, observations could also provide information on the vertical structure of ice hydrometeors. This possibility warrants further investigation, since it would expand the already powerful dataset offered by ICI.

Eriksson et al. (2020) estimated that ICI will deliver similar accuracy to the CloudSat- and CALIPSO-based DARDAR product within the sensitivity range of IWP, and we keep this opinion. Good retrieval accuracy was maintained for Z_m and D_m, even with the more diverse set of particle models used. In regard to Z_m, ICI cannot compete with radars. However, we have shown that it is possible to retrieve D_m from ICI observations. Given that ICI spans microwave and sub-millimetre

795 wavelengths with multiple channels, it is possible that there is more information available to constrain D_m, allowing ICI to contend with radars.

Notably, both the CloudSat and CALIPSO satellites ended operations in 2023. Since alternative data sources on atmospheric ice mass suffer from inconsistencies and uncertainties, we are now left with a significant observational gap. Yet, the future is promising; the launch of EarthCare is expected in 2024 (Illingworth et al., 2015) and ICI will be launched in the following years.

800 EarthCare will offer radar-lidar cloud profile measurements, acting as a continuation of the CloudSat and CALIPSO pairing. As this study has demonstrated, ICI that will offer observations specially suited to the measurement of ice hydrometeors. With these new sources of data, we move closer to achieving the reliable and consistent global datasets of ice mass that we need.

Appendix A: Neural network architecture and input data

The architecture of the QRNN used to perform the retrievals in this study is as follows:

805
– A single network was trained to predict all outputs IWP, IWC, Z_m, D_m. The network consisted of multiple fully-connected layers for all outputs, with several final layers for each output individually.

– The QRNN was trained to predict 99 uniformly spaced quantile levels $\tau \in [0.01, 0.99]$ for all output variables.

– Hyper-parameters, such as number of neurons, number of layers, and batch size were selected after training multiple networks and subsequently comparing their performance on the test dataset.

810 The training and validation data was structured and processed as follows:

– The input data consisted of antenna-weighted brightness temperatures for each of the 13 ICI channels. A classification of surface type, surface temperature, and surface pressure (included as variables in the retrieval database) were included as ancillary input data.

– Uncertainty due to thermal noise is generated as a Gaussian with zero mean and variance $\mathrm{NE\Delta T}_j^2$, where j is the ICI
815 channel. To avoid overfitting, noise was randomly generated and added to each batch of training data, during each epoch of training. Also in each training epoch and batch of training data, 1% of the surface types were shuffled, both to avoid overfitting of the model and to safeguard against any misclassification of surface type in the dataset.

– Due to the large range of magnitudes possible for IWP, a log-linear transformation was applied, where all values of IWP less than 1.0 kg m^{-2} were transformed into logarithmic space. Additionally, any cases of IWP less than 10^{-4} kg
820 m^{-2} were replaced with a value uniformly sampled from the range $[10^{-6}, 10^{-4}]$ and subsequently transformed in to logarithmic space. The remainder of the data were linearly normalised according to the statistics of the training set.

Table B1. Definitions of the MARSS and ISMAR channels simulated in this study. Not all channel IDs are official, and are used only in this study to simplify their representation.

Sensor	Channel Frequency (GHz)	Polarisation	Channel ID
MARSS	157.07 ± 2.60	H	M-17-H
MARSS	183.25 ± 3.00	H	M-19-H
MARSS	183.25 ± 7.00	H	M-20-H
ISMAR	243.20 ± 2.50	H	I-1-H
ISMAR	243.20 ± 2.50	V	I-1-V
ISMAR	325.15 ± 1.50	V	I-2-V
ISMAR	325.15 ± 3.50	V	I-3-V
ISMAR	325.15 ± 9.50	V	I-4-V
ISMAR	664.00 ± 4.20	H	I-5-H
ISMAR	664.00 ± 4.20	V	I-5-V

Appendix B: ISMAR and MARSS channels

Rather than refer to each ISMAR and MARSS channel by its specific frequency in Sect. 5.1.2, IDs were assigned to each channel. The channel IDs are presented in Table B1. The numbering of the MARSS channel IDs corresponds to the official IDs

825 (McGrath and Hewison, 2001). The channel IDs chosen for ISMAR channels are not official. The numbering of the ISMAR channels simply follows the order in which they appear in the table, with the same number applied to channels at the same frequency and bandwidth.

Appendix C: Cloud signal-based retrievals

Retrievals were performed using a model trained with simulated ICI T_a as the primary input. In contrast, the operational L2

830 product at EUMETSAT will perform inversions of ΔT_a. Our decision to focus on T_a is driven by the desire to minimise the complexity of the retrieval problem.

However, given that ΔT_a is a key component in the retrieval algorithm, we wished to test the use of ΔT_a as a model input in order to check that there was no major disadvantage arising from the choice to use T_a. We continued to exclude the pre-processing steps performed in the algorithm, thus performing a direct comparison between T_a and ΔT_a when used as retrieval

835 model inputs.

A QRNN was trained on the same training and validation sets as for the original model. The model was trained using the same ancillary inputs (see Appendix A), but the primary input changed to ΔT_a values for each ICI channel. Inversions were
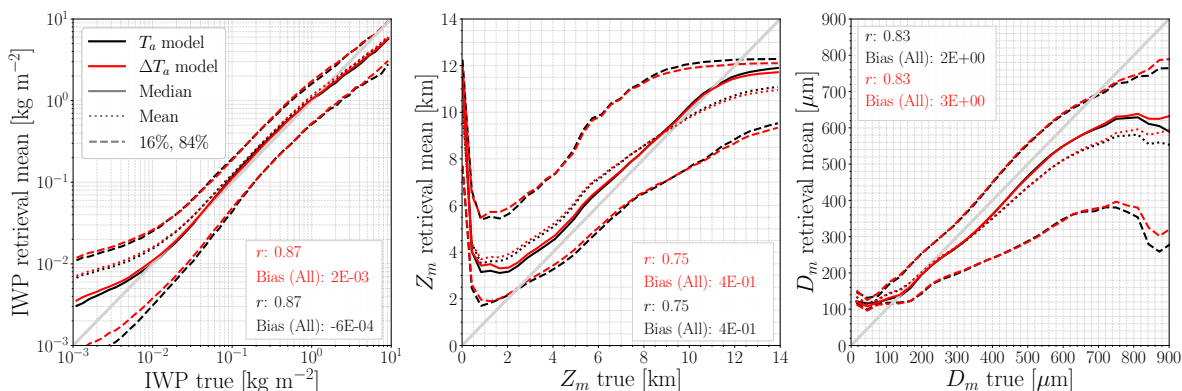
Figure C1. Retrieval performance for IWP, Z_m and D_m. The predictions from the T_a model, shown in black, are the same as shown in Fig. 9. The predictions from the ΔT_a model, shown in red, are a result of training a model to predict given a vector of ΔT_a, instead of T_a. The ancillary input data and outputs were kept the same in both models. r indicates the correlation coefficient. Bias is given in units of kg m^{-2}, km, and µm for IWP, Z_m and D_m, respectively.

performed on the same test data partition as originally used. In other words, ΔT_a are inverted to obtain IWP, Z_m, and D_m, where the 'true' values of IWP, Z_m, and D_m are the same for both retrieval models.

840   Both models exhibited very similar behaviour when predicting all three variables (Fig. C1). The most obvious differences occur at the extreme ends of the shown ranges. Both models produce the same correlation coefficient (at the precision shown), and this behaviour is again seen across all three variables. In the case of IWP, a bias of -6×10^{-4} kg m^{-2} is seen for the ΔT_a model. This is marginally smaller than the bias of 2×10^{-3} kg m^{-2} of the T_a model, and notably negative rather than positive. The bias of the D_m ΔT_a predictions is slightly higher than for the T_a model, at 3 µm as opposed to 2 µm.

845   The consistency in performance between the two models suggests that the all-sky simulations T_a appear equally capable as ΔT_a in the context of machine learning retrievals. However, it is stressed that this conclusion is made in the context of machine learning retrievals, and cannot necessarily be extended to the BMCI method used for the L2 product.

Explaining the minor discrepancies between the models is somewhat challenging. The bias was computed across the entire test dataset and not limited to the dataset represented in Fig. C1. It is therefore likely that the difference in bias occurs in 850 regions where the model is less well-trained. This is supported by the fact that the inversions produce almost identical results in regions where we expect the model to be well-trained, such as above the sensitivity threshold of ICI and where there exists a large amount of data (e.g. 10^{-2} kg m$^{-2} \leq$ IWP ≤ 1 kg m^{-2}). Given the small scale of the differences, the differences could be attributed to the nature of neural networks. Although the two networks were similarly configured, the learning process and final weights will not be identical, thus accounting for the small variation between models.

855 *Code and data availability.* The code used for analysis and plotting is available at https://doi.org/10.5281/zenodo.10839090 (May, 2024). The data pertaining to ICI used in this study are available under license for non-commercial purposes and on the condition of no redistribution by contacting EUMETSAT (vinia.mattioli@eumetsat.int). DARDAR- cloud v3-10 data are available at https://www.icare.univ-lille.fr/dardar/data-access/ (last access: 19 March 2024). GPM GMI data were accessed via NASA's Goddard Earth Services Data and Informa- tion Services Center (GES DISC), at https://gpm.nasa.gov/ (last access: 19 March 2024).

860 *Author contributions.* All authors contributed to the development of the database generation framework. PE provided input on all results and contributed to revisions of the paper. EM led the writing of the manuscript, and performed the visualisation and analysis of the results, with the exception of the GMI simulations which were visualised and analysed by HH. BE performed the ISMAR simulations and HH performed the GMI simulations.

Competing interests. The authors declare that they have no conflict of interest.

865 *Acknowledgements.* The development of the retrieval database was performed as part of the EUMETSAT study "Development of a cloud radiation database for EPS-SG ICI". Retrievals were performed using the QRNN package developed and made available by Simon Pfreundschuh, available at https://github.com/simonpf/quantnn, (last access: 22 February 2024).
The authors would also like to thank both Simon Pfreundschuh and Adrià Amell for their input and guidance on machine learning.
Computations were performed using resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the
870 Swedish National Infrastructure for Computing (SNIC).

Financial support. The research was partly funded by the Swedish National Space agency (grant no. 2021-00077). Development of the database was performed under the EUMETSAT study "Development of a cloud radiation database for EPS-SG ICI", contract EUM/-CO/21/4600002601/VM.

References

875   Accadia, C., Mattioli, V., Colucci, P., Schlüssel, P., D'Addio, S., Klein, U., Wehr, T., and Donlon, C.: Microwave and Sub-mm Wave
Sensors: A European Perspective, in: Satellite Precipitation Measurement: Volume 1, edited by Levizzani, V., Kidd, C., Kirschbaum, D. B.,
Kummerow, C. D., Nakamura, K., and Turk, F. J., Advances in Global Change Research, pp. 83–97, Springer International Publishing,
Cham, https://doi.org/10.1007/978-3-030-24568-9_5, 2020.

Amell, A., Eriksson, P., and Pfreundschuh, S.: Ice water path retrievals from Meteosat-9 using quantile regression neural networks, Atmos.
880   Meas. Tech., 15, 5701–5717, https://doi.org/10.5194/amt-15-5701-2022, 2022.

Barker, H. W., Jerg, M. P., Wehr, T., Kato, S., Donovan, D. P., and Hogan, R. J.: A 3D cloud-construction algorithm for the EarthCARE
satellite mission, Q. J. Roy. Meteor. Soc., 137, 1042–1058, https://doi.org/10.1002/qj.824, 2011.

Barlakas, V. and Eriksson, P.: Three dimensional radiative effects in passive millimeter/sub-millimeter all-sky observations, Remote Sensing,
12, 531, https://doi.org/10.3390/rs12030531, 2020.

885   Barlakas, V., Geer, A. J., and Eriksson, P.: Introducing hydrometeor orientation into all-sky microwave and submillimeter assimilation,
Atmos. Meas. Tech., 14, 3427–3447, https://doi.org/10.5194/amt-14-3427-2021, 2021.

Barlakas, V., Galligani, V. S., Geer, A. J., and Eriksson, P.: On the accuracy of RTTOV-SCATT for radiative transfer at all-sky microwave
and submillimeter frequencies, J. Quant. Spectrosc. Radiat. Transfer, 283, 108 137, https://doi.org/10.1016/j.jqsrt.2022.108137, 2022.

Bolot, M., Harris, L. M., Cheng, K.-Y., Merlis, T. M., Blossey, P. N., Bretherton, C. S., Clark, S. K., Kaltenbaugh, A., Zhou, L., and
890   Fueglistaler, S.: Kilometer-scale global warming simulations and active sensors reveal changes in tropical deep convection, npj Clim.
Atmos. Sci., 6, 1–8, https://doi.org/10.1038/s41612-023-00525-w, number: 1 Publisher: Nature Publishing Group, 2023.

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma,
A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, Nat. Geosci., 8, 261–268,
https://doi.org/10.1038/ngeo2398, 2015.

895   Brath, M., Fox, S., Eriksson, P., Harlow, R. C., Burgdorf, M., and Buehler, S. A.: Retrieval of an ice water path over the ocean from ISMAR
and MARSS millimeter and submillimeter brightness temperatures, Atmos. Meas. Tech., 11, 611–632, https://doi.org/10.5194/amt-11-
611-2018, 2018.

Brath, M., Ekelund, R., Eriksson, P., Lemke, O., and Buehler, S. A.: Microwave and submillimeter wave scattering of oriented ice particles,
Atmos. Meas. Tech., 13, 2309–2333, https://doi.org/10.5194/amt-13-2309-2020, 2020.

900   Buehler, S. A., Jiménez, C., Evans, K. F., Eriksson, P., Rydberg, B., Heymsfield, A. J., Stubenrauch, C. J., Lohmann, U., Emde, C., John,
V. O., Sreerekha, T. R., and Davis, C. P.: A concept for a satellite mission to measure cloud ice water path, ice particle size, and cloud
altitude, Q. J. R. Meteorol. Soc., 133, 109–128, https://doi.org/10.1002/qj.143, 2007.

Buehler, S. A., Mendrok, J., Eriksson, P., Perrin, A., Larsson, R., and Lemke, O.: ARTS, the Atmospheric Radiative Transfer Simulator –
version 2.2, the planetary toolbox edition, Geosci. Model Dev., 11, 1537–1556, https://doi.org/10.5194/gmd-11-1537-2018, 2018.

905   Cazenave, Q., Ceccaldi, M., Delanoë, J., Pelon, J., Groß, S., and Heymsfield, A.: Evolution of DARDAR-CLOUD ice cloud retrievals: new
parameters and impacts on the retrieved microphysical properties, Atmos. Meas. Tech., 12, 2819–2835, https://doi.org/10.5194/amt-12-
2819-2019, 2019.

Davis, C. P., Evans, K. F., Buehler, S. A., Wu, D. L., and Pumphrey, H. C.: 3-D polarised simulations of space-borne passive mm/sub-mm
midlatitude cirrus observations: a case study, Atmos. Chem. Phys., 7, 4149–4158, https://doi.org/https://doi.org/10.5194/acp-7-4149-2007,
910   2007.

Defer, E., Galligani, V. S., Prigent, C., and Jimenez, C.: First observations of polarized scattering over ice clouds at close-to-millimeter wavelengths (157 GHz) with MADRAS on board the Megha-Tropiques mission, J. Geophys. Res. Atmos., 119, 12,301–12,316, https://doi.org/10.1002/2014JD022353, 2014.

Delanoë, J. M. E., Heymsfield, A. J., Protat, A., Bansemer, A., and Hogan, R. J.: Normalized particle size distribution for remote sensing application, J. Geophys. Res. Atmos., 119, 4204–4227, https://doi.org/10.1002/2013JD020700, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JD020700, 2014.

Dong, P., Liu, L., Li, S., Husi, L., Hu, S., and Bu, L.: A novel ice cloud retrieval algorithm for submillimeter wave radiometers: simulations and application to an airborne experiment, IEEE T. Geosci. Remote, 61, 1–9, https://doi.org/10.1109/TGRS.2023.3274633, 2023.

Duncan, D. I. and Eriksson, P.: An update on global atmospheric ice estimates from satellite observations and reanalyses, Atmos. Chem. Phys., 18, 11 205–11 219, https://doi.org/10.5194/acp-18-11205-2018, 2018.

Ekelund, R. and Eriksson, P.: Impact of ice aggregate parameters on microwave and sub-millimetre scattering properties, J. Quant. Spectrosc. Radiat. Transfer, 224, 233–246, https://doi.org/10.1016/j.jqsrt.2018.11.013, 2019.

Ekelund, R., Eriksson, P., and Pfreundschuh, S.: Using passive and active observations at microwave and sub-millimetre wavelengths to constrain ice particle models, Atmos. Meas. Tech., 13, 501–520, https://doi.org/10.5194/amt-13-501-2020, 2020.

Eriksson, P., Rydberg, B., Sagawa, H., Johnston, M. S., and Kasai, Y.: Overview and sample applications of SMILES and Odin-SMR retrievals of upper tropospheric humidity and cloud ice mass, Atmos. Chem. Phys., 14, 12 613–12 629, https://doi.org/10.5194/acp-14-12613-2014, 2014.

Eriksson, P., Ekelund, R., Mendrok, J., Brath, M., Lemke, O., and Buehler, S. A.: A general database of hydrometeor single scattering properties at microwave and sub-millimetre wavelengths, Earth Syst. Sci. Data, 10, 1301–1326, https://doi.org/10.5194/essd-10-1301-2018, 2018.

Eriksson, P., Rydberg, B., Mattioli, V., Thoss, A., Accadia, C., Klein, U., and Buehler, S. A.: Towards an operational Ice Cloud Imager (ICI) retrieval product, Atmos. Meas. Tech., 13, 53–71, https://doi.org/10.5194/amt-13-53-2020, 2020.

Evans, K. F., Walter, S. J., Heymsfield, A. J., and McFarquhar, G. M.: Submillimeter-wave cloud ice radiometer: simulations of retrieval algorithm performance, J. Geophys. Res. Atmos., 107, AAC 2–1–AAC 2–21, https://doi.org/10.1029/2001JD000709, 2002.

Evans, K. F., Wang, J. R., O'C Starr, D., Heymsfield, G., Li, L., Tian, L., Lawson, R. P., Heymsfield, A. J., and Bansemer, A.: Ice hydrometeor profile retrieval algorithm for high-frequency microwave radiometers: application to the CoSSIR instrument during TC4, Atmos. Meas. Tech., 5, 2277–2306, https://doi.org/10.5194/amt-5-2277-2012, 2012.

Field, P. R., Heymsfield, A. J., and Bansemer, A.: Snow size distribution parameterization for midlatitude and tropical ice clouds, J. Atmos. Sci., 64, 4346–4365, https://doi.org/10.1175/2007JAS2344.1, 2007.

Fox, S., Lee, C., Moyna, B., Philipp, M., Rule, I., Rogers, S., King, R., Oldfield, M., Rea, S., Henry, M., Wang, H., and Harlow, R. C.: ISMAR: an airborne submillimetre radiometer, Atmos. Meas. Tech., 10, 477–490, https://doi.org/10.5194/amt-10-477-2017, 2017.

Fox, S., Mattioli, V., Turner, E., Vance, A., Cimini, D., and Gallucci, D.: An evaluation of atmospheric absorption models at millimetre and sub-millimetre wavelengths using airborne observations, EGUsphere, pp. 1–28, https://doi.org/10.5194/egusphere-2024-229, 2024.

Geer, A. J., Bauer, P., Lonitz, K., Barlakas, V., Eriksson, P., Mendrok, J., Doherty, A., Hocking, J., and Chambon, P.: Bulk hydrometeor optical properties for microwave and sub-millimetre radiative transfer in RTTOV-SCATT v13.0, Geosci. Model Dev., 14, 7497–7526, https://doi.org/10.5194/gmd-14-7497-2021, 2021.

Gong, J. and Wu, D. L.: Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements, Atmos. Chem. Phys., 17, 2741–2757, https://doi.org/10.5194/acp-17-2741-2017, 2017.

Harlow, C.: Millimeter microwave emissivities and effective temperatures of snow-covered surfaces: evidence for lambertian surface scatter-
950   ing, IEEE T. Geosci. Remote, 47, 1957–1970, https://doi.org/10.1109/TGRS.2008.2011893, 2009.

Harlow, C. and Essery, R.: Tundra snow emissivities at MHS frequencies: MEMLS validation using airborne microwave data measured
during CLPX-II, IEEE. T. Geosci. Remote., 50, 4262–4278, https://doi.org/10.1109/TGRS.2012.2193132, 2012.

Hewison, T., Selbach, N., Heygster, G., Taylor, J., and Mcgrath, A.: Airborne measurements of Arctic sea ice, glacier and snow emissivity at
24-183 GHz, vol. 5, pp. 2851–2855, Toronto, ON, Canada, https://doi.org/10.1109/IGARSS.2002.1026797, 2002.

955   Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D. P.,
Fukuda, S., Hirakata, M., Hogan, R. J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T. Y., Nishizawa, T., Ohno, Y.,
Okamoto, H., Oki, R., Sato, K., Satoh, M., Shephard, M. W., Velázquez-Blázquez, A., Wandinger, U., Wehr, T., and Zadelhoff, G.-J. v.:
The EarthCARE satellite: the next step forward in global measurements of clouds, aerosols, precipitation, and radiation, Bull. Amer. Met.
Soc., 96, 1311–1332, https://doi.org/10.1175/BAMS-D-12-00227.1, 2015.

960   Kanngiesser, F. and Eriksson, P.: Cautious note on using the discrete dipole approximation for inhomogeneous, spherical scatterers with high
optical contrast, Opt. Lett., 47, 4203–4206, https://doi.org/10.1364/OL.463561, 2022.

Kaur, I., Eriksson, P., Pfreundschuh, S., and Duncan, D. I.: Can machine learning correct microwave humidity radiances for the influence of
clouds?, Atmos. Meas. Tech., 14, 2957–2979, https://doi.org/10.5194/amt-14-2957-2021, 2021.

Kaur, I., Eriksson, P., Barlakas, V., Pfreundschuh, S., and Fox, S.: Fast radiative transfer approximating ice hydrometeor orientation and its
965   implication on IWP retrievals, Remote Sensing, 14, 1594, https://doi.org/10.3390/rs14071594, 2022.

Kim, J., Shin, D.-B., and Kim, D.: Effects of inhomogeneous ice particle habit distribution on passive microwave radiative transfer simula-
tions, IEEE T. Geosci. Remote, 62, 1–20, https://doi.org/10.1109/TGRS.2024.3349484, 2024.

Kulie, M. S., Bennartz, R., Greenwald, T. J., Chen, Y., and Weng, F.: Uncertainties in microwave properties of frozen precipitation: implica-
tions for remote sensing and data assimilation, J. Atmos. Sci., 67, 3471–3487, https://doi.org/10.1175/2010JAS3520.1, 2010.

970   Liebe, H.: Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz, in: AGARD Conference
Proceedings 542, Palma de Mallorca, Spain, 1993.

Mattioli, V., Accadia, C., Ackermann, J., Di Michele, S., Hans, I., Schlüssel, P., Colucci, P., and Canestri, A.: The EUMETSAT Polar System
- Second Generation (EPS-SG) Passive Microwave and Sub-mm Wave Missions, in: PhotonIcs & Electromagnetics Research Symposium,
pp. 3926–3933, https://doi.org/10.1109/PIERS-Spring46901.2019.9017822, 2019.

975   May, E.: The Ice Cloud Imager: retrieval of frozen water column properties - Code, https://doi.org/10.5281/zenodo.10839090, 2024.

McGrath, A. and Hewison, T.: Measuring the accuracy of MARSS—An airborne microwave radiometer, J. Atmos. Ocean. Tech., 18, 2003–
2012, https://doi.org/10.1175/1520-0426(2001)018<2003:MTAOMA>2.0.CO;2, 2001.

Mlawer, E. J., Payne, V. H., Moncet, J.-L., Delamere, J. S., Alvarado, M. J., and Tobin, D. C.: Development and recent evaluation of the
MT_CKD model of continuum absorption, Philos. T. R. Soc. A, 370, 2520–2556, https://doi.org/10.1098/rsta.2011.0295, 2012.

980   Moradi, I., Stegmann, P., Johnson, B., Barlakas, V., Eriksson, P., Geer, A., Gelaro, R., Kalluri, S., Kleist, D., Liu, Q., and Mccarty, W.:
Implementation of a discrete dipole approximation scattering database Into community radiative transfer model, J. Geophys. Res. Atmos.,
127, https://doi.org/10.1029/2022JD036957, 2022.

Munchak, S. J., Ringerud, S., Brucker, L., You, Y., De Gelis, I., and Prigent, C.: An active–passive microwave land surface database from
GPM, IEEE T. Geosci. Remote, 58, 6224–6242, https://doi.org/10.1109/TGRS.2020.2975477, 2020.

985 O'Shea, S. J., Choularton, T. W., Lloyd, G., Crosier, J., Bower, K. N., Gallagher, M., Abel, S. J., Cotton, R. J., Brown, P. R. A., Fugal, J. P., Schlenczek, O., Borrmann, S., and Pickering, J. C.: Airborne observations of the microphysical structure of two contrasting cirrus clouds, J. Geophys. Res. Atmos., 121, 13,510–13,536, https://doi.org/10.1002/2016JD025278, 2016.

Pfreundschuh, S.: Improving satellite measurements of clouds and precipitation using machine learning, Ph.D. thesis, Chalmers University of Technology, https://research.chalmers.se/en/publication/530164, iSBN: 9789179056575, 2022.

990 Pfreundschuh, S., Eriksson, P., Duncan, D., Rydberg, B., Håkansson, N., and Thoss, A.: A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems, Atmos. Meas. Tech., 11, 4627–4643, https://doi.org/10.5194/amt-11-4627-2018, 2018.

Pfreundschuh, S., Eriksson, P., Buehler, S. A., Brath, M., Duncan, D., Larsson, R., and Ekelund, R.: Synergistic radar and radiometer retrievals of ice hydrometeors, Atmos. Meas. Tech., 13, 4219–4245, https://doi.org/10.5194/amt-13-4219-2020, 2020.

Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., and Norrestad, T.: GPROF-NN: a neural-network-based implementation of

995 the Goddard Profiling Algorithm, Atmos. Meas. Tech., 15, 5033–5060, https://doi.org/10.5194/amt-15-5033-2022, 2022a.

Pfreundschuh, S., Fox, S., Eriksson, P., Duncan, D., Buehler, S. A., Brath, M., Cotton, R., and Ewald, F.: Synergistic radar and sub-millimeter radiometer retrievals of ice hydrometeors in mid-latitude frontal cloud systems, Atmos. Meas. Tech., 15, 677–699, https://doi.org/10.5194/amt-15-677-2022, 2022b.

Pickett, H. M., Poynter, R. L., Cohen, E. A., Delitsky, M. L., Pearson, J. C., and Müller, H. S. P.: Submillimeter, millimeter, and microwave

1000 spectral line catalog, J. Quant. Spectrosc. Radiat. Transfer, 60, 883–890, https://doi.org/10.1016/S0022-4073(98)00091-0, 1998.

Prigent, C., Aires, F., Wang, D., Fox, S., and Harlow, C.: Sea-surface emissivity parametrization from microwaves to millimetre waves, Q. J. Roy. Meteor. Soc., 143, 596–605, https://doi.org/10.1002/qj.2953, 2017.

Risse, N.: Microwave emissivity of sea ice from airborne observations, Master's thesis, University of Bonn, 2021.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2 of *Series on Atmospheric, Oceanic and Planetary*

1005 *Physics*, World Scientific, https://doi.org/10.1142/3171, 2000.

Rosenkranz, P.: Absorption of microwaves by atmospheric gases, in: Atmospheric remote sensing by microwave radiometry, John Wiley and Sons, https://dspace.mit.edu/handle/1721.1/68611, 1993.

Rydberg, B.: EPS-SG ICI ice water path product: ATBD, https://research.chalmers.se/en/publication/514522, 2018.

Rydberg, B., Eriksson, P., and Buehler, S. A.: Prediction of cloud ice signatures in submillimetre emission spectra by means of ground-based

1010 radar and in situ microphysical data, Q. J. R. Meteorol. Soc., 133, 151–162, https://doi.org/10.1002/qj.151, 2007.

Rydberg, B., Eriksson, P., Buehler, S. A., and Murtagh, D. P.: Non-Gaussian Bayesian retrieval of tropical upper tropospheric cloud ice and water vapour from Odin-SMR measurements, Atmos. Meas. Tech., 2, 621–637, https://doi.org/10.5194/amt-2-621-2009, 2009.

Rydberg, B., May, E., Hallborn, H., Eriksson, P., and Kaur, I.: Development of a cloud radiation database for EPS-SG ICI, Tech. rep., Chalmers University of Technology, https://research.chalmers.se/en/publication/535268, 2023.

1015 Stephens, G. L., Tsay, S.-C., Stackhouse, P. W., and Flatau, P. J.: The relevance of the microphysical and radiative properties of cirrus clouds to climate and climatic feedback, J. Atmos. Sci., 47, 1742–1754, https://doi.org/10.1175/1520-0469(1990)047<1742:TROTMA>2.0.CO;2, 1990.

Waliser, D. E., Li, J.-L. F., Woods, C. P., Austin, R. T., Bacmeister, J., Chern, J., Del Genio, A., Jiang, J. H., Kuang, Z., Meng, H., Minnis, P., Platnick, S., Rossow, W. B., Stephens, G. L., Sun-Mack, S., Tao, W.-K., Tompkins, A. M., Vane, D. G., Walker, C., and Wu, D.: Cloud ice:

1020 A climate model challenge with signs and expectations of progress, J. Geophys. Res. Atmos., 114, https://doi.org/10.1029/2008JD010015, 2009.

Wang, D., Prigent, C., Aires, F., and Jimenez, C.: A statistical retrieval of cloud parameters for the millimeter wave Ice Cloud Imager on board MetOp-SG, IEEE Access, PP, 1–1, https://doi.org/10.1109/ACCESS.2016.2625742, 2016.

Wang, D., Prigent, C., Kilic, L., Fox, S., Harlow, C., Jimenez, C., Aires, F., Grassotti, C., and Karbou, F.: Surface emissivity at microwaves to millimeter waves over polar regions: parameterization and evaluation with aircraft experiments, J. Atmos. Ocean. Tech., 34, 1039–1059, https://doi.org/10.1175/JTECH-D-16-0188.1, 2017.

Wu, D. L., Jiang, J. H., Read, W. G., Austin, R. T., Davis, C. P., Lambert, A., Stephens, G. L., Vane, D. G., and Waters, J. W.: Validation of the Aura MLS cloud ice water content measurements, J. Geophys. Res. Atmos., 113, https://doi.org/10.1029/2007JD008931, 2008.

Yau, M. K. and Rogers, R. R.: A short course in cloud physics, Third Edition, Butterworth-Heinemann, 1989.

1025